M

# Journal of Experimental Psychology

ARTHUR W. MELTON, Editor
HUMAN RESOURCES RESEARCH CENTER
LACKLAND AIR FORCE BASE
SAN ANTONIO, TEXAS

CONSULTING EDITORS

Judson S. Brown, *State University of Iowa*
Frank A. Geldard, *University of Virginia*
James J. Gibson, *Cornell University*
Clarence H. Graham, *Columbia University*
David A. Grant, *University of Wisconsin*
Lloyd G. Humphreys, *Human Resources Research Center*
Donald B. Lindsley, *University of California, Los Angeles*
Neal E. Miller, *Yale University*
Edwin B. Newman, *Harvard University*
Benton J. Underwood, *Northwestern University*
Delos D. Wickens, *Ohio State University*

Lorraine Bouthilet, Managing Editor

RECEIVED SEP 18 1952 PSYCH. ABSTR.

## CONTENTS

*Rigidity as Learned Behavior:* H. M. SCHRODER AND J. B. ROTTER ..... 141

*Assimilation in the Immediate Reproduction of Visually Perceived Figures:* J. S. BRUNER, R. D. BUSIEK, AND A. L. MINTURN 151

*The Effects of Failure and Retroactive Inhibition on Mediated Generalization:* B. B. MURDOCK, JR. 156

*Retroactive Inhibition in Recall and Recognition:* L. POSTMAN ..... 165

*Experimental Extinction As a Function of the Intertrial Intervals during Conditioning and Extinction:* W. H. TEICHNER ..... 170

*Shifts in Scale Values of Attitude Statements As a Function of the Composition of the Scale:* E. FEHRER ..... 179

*The Effect of Learning on the Predictability of Psychomotor Performance:* B. REYNOLDS 189

*An Experimental Analysis of the Errors in Rote Serial Learning:* J. DEESE AND F. H. KRESSE 199

*Effect of Intertrial Interval during Acquisition on Extinction of the Conditioned Eyelid Response following Partial Reinforcement:* D. A. GRANT, L. M. SCHIPPER, AND B. M. ROSS 203

*Comparisons of Forward, Simultaneous, Backward, and Pseudo-Conditioning:* M. E. FITZWATER AND M. N. REISMAN 211

*Human Discrimination Learning with Simultaneous and Successive Presentation of Stimuli:* H. B. LOESS AND C. P. DUNCAN 215

*Some Determinants of Rigidity in Discrimination-Reversal Learning:* A. H. BUSS ..... 222

American Psychological Association

Vol. 44 No. 3 September 1952

# JOURNAL OF EXPERIMENTAL PSYCHOLOGY

The JOURNAL OF EXPERIMENTAL PSYCHOLOGY is published monthly, two volumes per year, by the American Psychological Association, Inc. The subscription rate per volume is $7.00, or $14.00 annually. Single copies are $1.25. Subscriptions, orders, and business communications should be addressed to the American Psychological Association, Inc., 1515 Massachusetts Ave., N.W., Washington 5, D. C.

This JOURNAL publishes original experimental investigations which are intended to contribute toward the development of psychology as an experimental science. Studies with normal human subjects are favored over studies involving abnormal or animal subjects. Studies in applied experimental psychology or engineering psychology may be accepted if they have broad implications for experimental and theoretical psychology. Normally, articles of a length exceeding 15 printed pages cannot be accepted. However, an integrated series of studies accomplished simultaneously (e.g., most doctoral dissertations) must be presented in a single article, rather than in a series of articles.

Articles are published in the order of their receipt, except in rare circumstances. No free reprints are supplied to authors. All of the cost of an author's alterations in galley proof is charged to the author, and one-half of the cost of tables, cuts, and formulas. Priority in publication is given to articles whose authors assume the full cost of publication.

Address all articles submitted for regular or priority publication to the editor, Arthur W. Melton, Human Resources Research Center, Lackland Air Force Base, San Antonio, Texas.

**Manuscripts submitted to the JOURNAL *must* adhere to the conventions concerning reference citations, preparation of tables and figures, manuscript format, etc. as described in the Publication Manual of the American Psychological Association (*Psychol. Bull.*, 1952, 49 (4, Suppl.), 389–449. In particular, the organization of the manuscript should be text, references, footnotes, tables (one to a page), figure titles (typed on separate page), and figures. All typed material *must* be double-spaced with ample margins. The term "subject," with its various forms, is abbreviated *S*, *Ss*, *S*'s, and *Ss*'; analogous abbreviations are used for "experimenter" (*E*) and "observer" (*O*). When in doubt about the practices of the JOURNAL, authors should examine a recent issue.**

---

Entered as second-class matter, February 6, 1937, at the post office at Lancaster, Pa., under the Act of March 3, 1879.

Acceptance for mailing at the special rate of postage provided for in paragraph (d-2), Section 34.40, P. L. & R. of 1948, authorized September 11, 1947.

Journal of
Experimental Psychology

VOL. 43, NO. 3 SEPTEMBER, 1952

# RIGIDITY AS LEARNED BEHAVIOR

HAROLD M. SCHRODER AND JULIAN B. ROTTER

*Ohio State University*

Although psychologists have been concerned for some time with inflexible or fixated behavior, systematic study of the concept of rigidity has been relatively recent. Freud with his concept of fixation and Adler with his concept of style of life have described behavior which appears to be consistently inappropriate to a present set of cues or at least is responsive only to a very limited set of cues in a variety of situations. These clinical approaches, however, fail to clearly describe the conditions under which this behavior will occur except, perhaps, at a high level of generality.

For Lewin (**4**), rigidity was a central construct. It was defined as "the state of the boundaries between psychological regions." Kounin (**3**) has applied Lewin's rigidity construct to the feebleminded. Goldstein (**1**) has approached the problem from the point of view of mental set, and Luchins (**5**) has placed emphasis on the field conditions rather than on the individual.

In general these approaches fail to describe the specific learning or training conditions which will account for individual differences in "rigid" behavior. Nor do they provide theories broad enough to account for the similarity of behavior seen in so-called feebleminded, neurotic, brain-injured, and some normal individuals who appear to have in common a lack of flexibility, a resistance to change or shift, or a repetitiveness of the same behavior in what appear to be a variety of situations.

The present study tests some hypotheses drawn from a social-learning theory of behavior developed by Rotter (**9**) and his students. From this point of view we would ask the question: What sequence of learning experiences would make for greater or lesser potentiality to try out alternative solutions in a problem situation rather than a repeated use of the same behavior? Restated, we might ask under what conditions does an individual enter a problem situation with an expectancy that a single behavior will lead to reinforcement and under what conditions will he expect that there may be more than one way of reaching the goal, the best solution varying from situation to situation. We would hypothesize that the looking for alternative solutions or pathways to goals is itself a higher-level behavior which is reinforceable and may have differential strengths following varied training

sequences. It might be called nonrigid behavior—its absence, rigid behavior. Interestingly enough, in this formulation it is the nonrigid behavior which is acquired, and rigidity as a specific trait or entity which is itself modifiable is nonexistent. That is, "rigidity" is a failure to learn something rather than an inherent or original trait. Although not working with the concept of rigidity as such, Harlow (**2**) has been particularly successful in demonstrating the existence of higher-level learning skills in monkeys which make for flexibility and adaptability.

If the hypothesis stated above were to be correct it would be necessary to develop its implications before constructing a test. In order to develop this behavior of looking for alternatives it would be necessary to provide a learning sequence in which apparently similar problems would lead to solution or reinforcement in different ways and *S* would be forced to adopt different solutions. The more times in the series the correct solutions (which could easily be arrived at) shifted, the more *S* would approach each new problem by asking which solution is right and the sooner he would perceive that his previous solution was incorrect. The *S* without such training would take longer to make shifts to changes in the situation but once such shifts would occur, he would then use his new solution more consistently than the *S* who has more reinforcement for trying out alternatives. Since the design called for a task in which generalization effects from previous experience and from one solution to another were at a minimum, it was felt necessary to develop a new task, rather than use the technique developed by Luchins (**5**). The method described below was designed to test these various aspects of the main hypothesis.

## Experimental Design

*Materials.*—Sixty-six cards were constructed, 3.5 in. square and consisting of thick cardboard. Each card contained (*a*) color (blue, gold, or silver), (*b*) an outside form, and (*c*) an inside form. These figures were drawn on the cards with thick lines using black India ink and are shown in Fig. 1 and 2.

The task consisted of grouping. Each trial consisted of *S* grouping six cards into three pairs

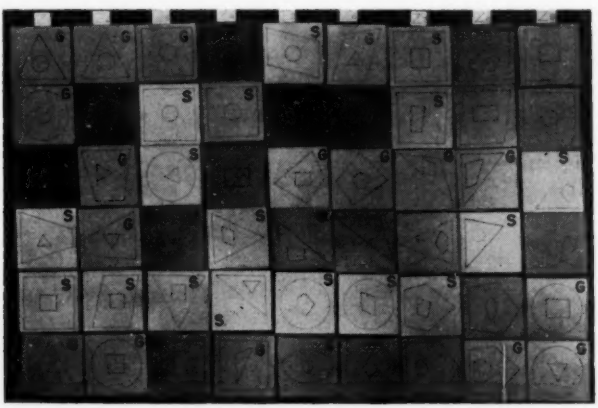

Fig. 1. Cards Used in Training Trials

Letters indicate the color of each card. All cards without letters are blue. G represents gold; S represents silver.

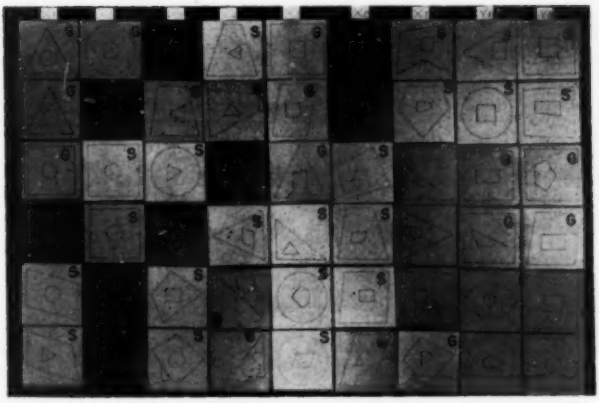

FIG. 2. Cards Used in Testing Trials

Letters indicate the color of each card. All cards without letters are blue. G represents gold; S represents silver.

on the basis of color, inside form, or outside form. Generally the cards were constructed so that grouping was only possible on one of the above bases. The experimental design is shown in Table 1. The table is read in terms of the following considerations: (*a*) Each letter (e.g., $x_1$) refers to a trial, i.e., when *S* is handed six cards which have to be grouped into three pairs on the basis of form or color. (*b*) $x$ = grouping possible on outside form only; $y$ = grouping possible on color only; $z$ = grouping possible on inside form only; $z'$ = grouping possible on inside and outside form. (*c*) No single grouping is difficult once the relevant principle is selected. (*d*) From $x_1$ to $x_7$ there is a progression from grouping on outside forms which are exactly alike to grouping on outside forms not so objectively similar. That is, the difficulty increases if difficulty is defined as greater objective dissimilarity of the forms which must be grouped together. (*e*) $x_8$ was given last in each training sequence, so that comparisons between the groups were possible. (*f*) $z_1$, $z_2$, and $z_3$ contained inside form as the basis of grouping into pairs. In each instance the forms to be paired were alike. (*g*) In $z_4$ the inside forms were the only basis for grouping, but were not objectively similar. (*h*) From $z'_5$ to $z'_9$ there is a progression in that (1) inside figures as a basis for grouping become more dissimilar, and (2) outside figures as a basis for grouping become more similar.

For Groups 1 and 2 the training sequences consist of training in grouping the cards on outside form for all trials. In Groups 3 and 4 the grouping was made on outside form and color. The experimental sequence, which is constant for each group, demands a new behavior or solution for all sequences on $z_4$ (inside form). This solution must be repeated four times, experimentally controlled so as to allow progressive predictions for each group. After this, *S* can change his basis of grouping back to outside form ($x$) on any trial. In this series ($z'_5$–$z'_9$) grouping on inside form becomes progressively more difficult and on outside form progressively less difficult.

TABLE 1

EXPERIMENTAL DESIGN

| Group | Training Trials 1 2 3 4 5 6 7 | Experimental Trials 1 2 3 4 5 6 7 8 9 |
|---|---|---|
| 1 | $x_1\ x_2\ x_3\ x_4\ x_6\ x_7\ x_5$ | $z_1\ z_2\ z_3\ z_4\ z'_5\ z'_6\ z'_7\ z'_8\ x_8$ |
| 2 | $x_1\ x_2\ x_3\ x_4\ x_5$ | |
| 3 | $x_1\ y_1\ y_2\ x_2\ x_3\ x_4\ x_5$ | |
| 4 | $x_1\ y_1\ x_2\ y_2\ x_3\ x_4\ x_5$ | |

The four training sequences are deductions from the hypothesis, Sequence 1 being the most rigid and Sequence 4 the most nonrigid in terms of expectancy for change to occur. In Sequence 2, *S* learns, apart from solving a problem by outside form, which all groups learn, to solve *all* problems by outside form. According to the hypothesis he is learning to expect a reinforcement by using the same behavior. In Sequence 3, *S* learns to expect a reinforcement from two behaviors and to expect a reinforcement by changing his solution on two trials. In Sequence 4, *S* learns the same behaviors as in Sequence 3, but reinforcement occurs after a change or shift in correct solution on four trials. Therefore, expectancy for a reinforcement to occur by changing solutions should be highest in Group 4.

Sequence 1 was included for two reasons: (*a*) The longer the training in building up an expectancy for reinforcement to follow a single behavior, the lower the expectancy that a change will be reinforced. (*b*) It provides a necessary control in that Sequences 3 and 4 have seven trials and it cannot be assumed that a longer training period, or sheer time in looking at the cards, has no effect on the results. Sequence 1 is expected to take an equivalent training time and Sequence 2, the lowest, having five trials.

The following quantitative measures may be made: (*a*) The time of grouping, in seconds, for each trial; (*b*) the trial in the z series ($z'_5 - x'_9$) where *S* changes his basis of grouping from z to *x*; (*c*) the number of times *S* changes from grouping on *x* to grouping on z, or vice versa, in the experimental series.

*Predictions.*—Four specific hypotheses or predictions were made regarding differences among the four groups.

*Hypothesis A.* There will be a progressive decrease in time taken to group on $z_1$ from Groups 1 to 4. This would be expected since this would be the reverse of the order in which the groups would stop trying to solve the task by an outside form (*x*) solution and would begin looking for alternative solutions.

*Hypothesis B.* There will be a progressive decrease in the difference scores between time taken to solve $z_1$ and $z_2$ from Groups 1 to 4. In Groups 3 and 4 the learning of z or inside form continues to increase the expectancy of change and the behavior of looking for alternatives. In Groups 1 and 2, although such an expectancy may arise, it is still relatively weak and these *S*s are likely to expect z or inside figures to be the correct solution now and to apply it immediately.

*Hypothesis C.* From Groups 1 to 4 the change back to grouping on *x* (outside form) in the z series will occur progressively earlier. In spite of the fact that Group 1 had the most *x* trials, Group 4 which has the highest potential for seeing alternative solutions should be the first to discover that it is possible to group on outside form in the z series.

*Hypothesis D.* There will be progressively more changes or alternations in z and *x* solutions moving from Groups 1 to 4. This hypothesis is an obvious corollary of the one above.

*Subjects.*—The *S*s consisted of 104 beginning psychology students who were randomly selected. The group contained about an equal number of males and females and they were fairly homogeneous in respect to age and intelligence. The *S*s were given the instructions without *E* knowing in advance to which group each *S* was assigned. This was decided immediately before giving the test by drawing a number from a pool so constructed that 26 *S*s would be assigned to each group.

*Instructions.*—When *S* entered the testing room, the cards were spread out on a table, face downward. The following instructions were read to each *S*:

"For the next 20 min. or so, I want you to do a series of simple tasks for me. As you can see I have a number of cards here, and at the moment they are turned face downward. On each trial, and there will be a number of trials, I will hand you six cards. Each card has a number of characteristics on it.

"Your task is to group these six cards into three pairs. It may become clear to you if I give you an example. Suppose I hand you six insects and ask you to group them into pairs. Insects have a great number of characteristics, e.g., wings, legs, feelers, etc.; the cards here are much simpler. However, you may try to group them on the number of legs they have, but since they all have six legs, no differentiation is possible. You may find three large and three small insects, but this is not a grouping in pairs. Eventually you may see that two have large wings, two medium-sized wings, and two have very small wings. This would be a legitimate grouping for our purposes because (*a*) all insects have wings—the characteristic chosen for grouping and (*b*) you only used one basis for grouping on that trial. For example, you would not mix the bases of grouping by putting two insects together because they had similar wings, two others because they had similar feelers, and two others together for some other reason.

"When six cards are grouped into pairs they must all have the characteristic chosen as the basis of grouping. Further, each pair must be alike in terms of this characteristic, and the three pairs must be differentiated in terms of the same characteristic.

"So the two cards you put together each time must be alike in some way. They need not be exactly alike, as long as they are near enough to be differentiated from the other pairs.

"Your task is to find the basis of grouping or the characteristic on the cards which will permit a grouping into three pairs, on each trial. You can group on *anything* you like, as long as you use a single basis which does separate them into three pairs. There is no right or wrong grouping, and no particular grouping is any better than another from my point of view. Just group them as quickly as you can each time I hand you six cards, but do not rush it. You will find the task fairly simple. Any questions?"

Few *S*s asked questions. If they did, *E* reread and elaborated the instructions, using the insect example. Every attempt was made to insure that each *S* had a clear notion of what was required and to stress that it was a free situation, in that he could use whatever bases he pleased. For *S* the only aim was to arrive at some kind of grouping.

The *S* was then instructed to: (*a*) sit with his back to the cards so that *E* could select each set of six cards without *S*'s knowledge and without disturbing him; (*b*) to start grouping immediately when *E* spread the cards out face up on the table in front of him; (*c*) when finished, the cards should be in three groups of two, separated for inspection and face upward; and (*d*) as soon as this was done he should notify *E*. This was stressed because it was pointed out that *E* had no other indication and could not check the time until he was notified.

The time of grouping was taken from the time *E* had the cards face up on the table in front of *S* to the time when *S* notified *E* he was finished. Apart from the time taken for each trial, *E* noted the basis of grouping on each trial in the z series.

## Results

The important results relating to the training ($x$) series are given in Table 2. Time in the $x$ series for Group 2, which had only five trials, was significantly less than the other groups but there are no differences approaching significance among Groups 1, 3, and 4. The reader may be struck by the disparity between Group 2 and the results obtained for all other groups over the training sequence. The differences seem larger than could could be expected on the basis of two extra trials. However, Groups 3 and 4 each have two groupings on color after establishing the form solution, while Group 1 contains two extra and more difficult outside form solutions to master. The difficulty of these form solutions was purposely increased to bring Group 1 up to a similar total training time as Groups 3 and 4. If Group 1, therefore, differs from Groups

TABLE 2

Mean Time (Seconds) for *Total Training Trials*, $X_5$ and All Testing Trials

| Trials | Group 1 | | Group 2 | | Group 3 | | Group 4 | |
|---|---|---|---|---|---|---|---|---|
| | Mean | *SD* | Mean | *SD* | Mean | *SD* | Mean | *SD* |
| | | | | *Training* | | | | |
| All | 211.3 | 85.2 | 69.0 | 27.1 | 196.2 | 62.0 | 209.8 | 106.0 |
| $x_5$ | 16.7 | 14.6 | 16.3 | 12.0 | 21.2 | 11.0 | 23.5 | 17.8 |
| | | | | *Testing* | | | | |
| $z_1$ | 24.00 | 11.80 | 19.77 | 10.42 | 15.42 | 8.77 | 11.59 | 4.42 |
| $z_2$ | 12.19 | 6.50 | 12.50 | 4.11 | 13.19 | 6.12 | 14.88 | 7.74 |
| $z_3$ | 12.80 | 8.08 | 13.80 | 9.14 | 13.54 | 8.60 | 16.14 | 6.72 |
| $z_4$ | 17.80 | 14.00 | 14.15 | 4.14 | 26.53 | 23.60 | 20.73 | 15.76 |
| $z'_5$ | 16.46 | 15.30 | 25.80 | 30.80 | 34.30 | 40.60 | 19.07 | 16.16 |
| $z'_6$ | 14.04 | 8.40 | 13.96 | 7.21 | 15.50 | 6.80 | 14.19 | 7.30 |
| $z'_7$ | 11.00 | 4.40 | 13.46 | 7.56 | 17.46 | 8.43 | 13.80 | 4.12 |
| $z'_8$ | 12.85 | 7.80 | 15.46 | 10.04 | 17.69 | 10.70 | 16.53 | 9.24 |
| $x_9$ | 13.81 | 6.10 | 18.34 | 14.50 | 14.76 | 7.18 | 16.34 | 10.40 |

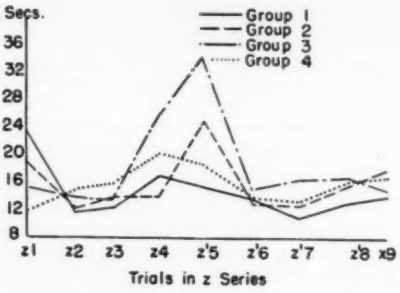


FIG. 3. Mean times of grouping on each trial in testing series for all groups

3 and 4 on any of our experimental measures it cannot be related to differences in training time.

It can be noted that there is an increasing progression in time of completing the last ($x_8$) task of the series for Groups 1 to 4. This suggests that differences in the groups due to more trying out of alternative pathways are already appearing in the last of the training series. Differences between any two groups, however, are not significant below the 10% level.

Trial by trial results for the experimental ($z$) series for all groups are shown in Table 2 and Fig. 3.

Bartlett's test for homogeneity of variance was applied to all measures relating to the four hypotheses. In all cases the assumption of homogeneity was supported. The $F$ ratios

TABLE 3

MEAN TRIAL ON WHICH EACH GROUP CHANGED FROM $z$ TO $x$ SOLUTIONS AND THE NUMBER OF ALTERNATIONS IN SOLUTION

| Group | Change from $z$ to $x$ | | Number of Changes | |
|---|---|---|---|---|
| | Mean | *SD* | Mean | *SD* |
| 1 | 7.423 | 1.590 | .6153 | .4800 |
| 2 | 7.230 | 1.440 | .8070 | .5500 |
| 3 | 6.5000 | 1.360 | 1.0000 | .4800 |
| 4 | 6.030 | 1.220 | 1.5760 | .6200 |

TABLE 4

$p$ LEVELS BETWEEN GROUPS FOR TRIAL $z_1$

| Groups | 2 | 3 | 4 |
|---|---|---|---|
| 1 | .15 | .01 | .01 |
| 2 | | .11 | .01 |
| 3 | | | .06 |

failed to reach the accepted criterion of significance on all individual test trials with the exception of Trial 1. Table 3 presents the means and standard deviations for (*a*) the mean trial at which each group changed from Solution $z$ (inside form) back to Solution $x$ (outside form) and (*b*) the mean number of alternations in the method of solution for each group in the testing series.

*Hypothesis A.*—An over-all analysis of variance for the four groups and time in solving the first experimental task ($z_1$) resulted in an $F$ of 7.66, significant at the 1% level. Individual probability levels are given in Table 4. It can be seen most clearly in Fig. 3, however, that a clear progression exists from Group 1 to Group 4 in the predicted direction.

*Hypothesis B.*—An analysis of variance of the differences between time for $z_1$ and $z_2$ was made to test Hypothesis B. Again the analysis showed significant differences between groups. An $F$ ratio of 11.75 was obtained, significant at the 1% level. Probability levels between groups are indicated in Table 5.

Inspection of Fig. 3 reveals a pro-

TABLE 5

$p$ LEVELS FOR DIFFERENCES BETWEEN TIME OF GROUPING $z_1$ AND $z_2$

| Groups | 2 | 3 | 4 |
|---|---|---|---|
| 1 | .01 | .01 | .01 |
| 2 | | .05 | .01 |
| 3 | | | .02 |

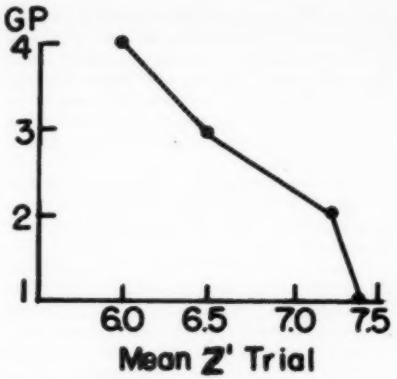


FIG. 4. Mean trial at which each group changed back to grouping on $x$ (outside form)

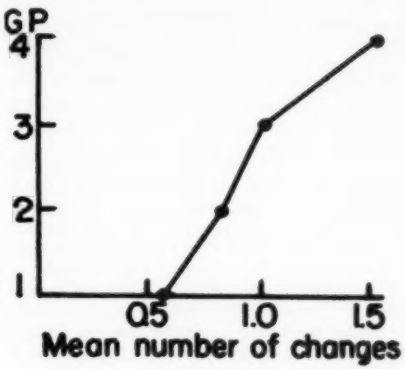


FIG. 5. Mean number of alternations in solution for each group in testing series

gression from Group 1 to Group 4 in the predicted direction. While the first two groups reduced their solution time considerably, Group 3 reduced only slightly and Group 4 took longer on $z_2$ than $z_1$. This tendency, particularly in relation to Group 4, is hard to account for on the basis of any hypothesis other than that the successful solution of $z_1$ strengthened or reinforced a behavior of looking for new or different solutions to the point that it seriously competed with a tendency to resolve $z_2$ with the same kind of behavior as $z_1$.

*Hypothesis C.*—To test Hypothesis C the mean trial in which the groups switched back to their first $x$ solution in the ambiguous trials $z'_5$ to $z'_8$ was calculated (Table 3). On these trials solution either by an $x$ hypothesis (outside form) or a $z$ hypothesis (inside form) is possible but is progressively more difficult for $z$ (inside form). Analysis of variance results in an $F$ ratio of 5.07, significant at the 1% level. Differences between groups are given in Table 6, while Fig. 4 shows a clear progression from Groups 1 to 4 in the predicted direction.

*Hypothesis D.*—Hypothesis D was tested by comparing the mean number of times that the groups changed solutions (or switched back and forth between inside form and outside form) in the ambiguous trials $z'_5$ through $z'_8$. Analysis of variance yielded an $F$ ratio of 14.66, significant at the 1% level. The progression along predicted lines for Groups 1 to 4 is shown in Fig. 5. Probability levels between groups are given in Table 7.

TABLE 6

$p$ LEVELS BETWEEN GROUPS FOR CHANGE FROM $z$ TO $x$ SOLUTIONS

| Groups | 2 | 3 | 4 |
|---|---|---|---|
| 1 | .65 | .05 | .01 |
| 2 | | .05 | .01 |
| 3 | | | .22 |

TABLE 7

$p$ LEVELS FOR NUMBER OF ALTERNATIONS IN SOLUTION

| Groups | 2 | 3 | 4 |
|---|---|---|---|
| 1 | .20 | .01 | .01 |
| 2 | | .20 | .01 |
| 3 | | | .01 |

## Discussion

All four specific hypotheses appear to be unequivocally supported by the results. In all the tests made, Groups 1, 2, 3, and 4 followed the same progressive order—Group 1 behaving in the most "rigid" manner and Group 4 in the least rigid manner.

Rigid behavior as here described is characterized as the approaching of a given situation with restricted attention to a given set of cues and an expectancy that there is a single pathway to reinforcement or solution that does not change. Such behavior makes for speed of performance and efficiency once the solution is learned, if the situation does not change. Nonrigid behavior, on the other hand, is characterized as the learning to attend to many cues and to expect that more than one pathway to reinforcement is present in a situation, or that the best or correct solution may be expected to be different or change in repeated trials. Such behavior makes for speed in grasping new principles or new solutions, for more alternation of behavior when more than one solution is possible, and for slightly less efficient performance when the situation remains the same and *S* presumably attends to cues not necessary for solution.

Conditions in this experiment leading to rigid behavior consist of repetition of the same or similar situations with reinforcement always following the same type of solution (outside form). The more trials the greater the rigidity. Nonrigid behavior was learned by introducing trials where an alternative solution was possible and sufficiently easy that it could be readily grasped when *S* discovered that the previously correct method would not work. The more times *S* is forced to adopt another solution the greater is the potential that he will expect and look for change in the situation.

By including Groups 2 and 3 as controls it was possible to demonstrate that the differences between Groups 1 and 4 are not merely a function of the number of trials in which outside form is the correct solution since Groups 3 and 4 were consistently less rigid than Group 2 as well as less rigid than 1. The inclusion of Group 3 made it possible to demonstrate that the results cannot be accounted for merely on the basis of the number of times color was the correct basis for grouping or on the basis of any positive transfer or generalization from color to inside form as a means of solution, since Group 4 with the same number of color solutions but more shifts or changes in the situation than Group 3 was consistently less rigid on all measures.

Reed (7) has obtained somewhat similar results in a study using the white rat. He found that longer training times on an initial problem resulted in slower subsequent learning when the correct solution was reversed. His results are comparable to those obtained with Group 1 which was slow on the first *z* solution but then improved with great rapidity.

These results are also consistent with the findings of Harlow (2) and his evidence of higher-level skills in the discrimination learning behavior of monkeys.

It seems to the authors that any explanation of the results in terms of more specific habits or skills cannot be logically maintained. Although it might be possible to explain the results of the first test of speed of grouping on the first *z* trial on the basis of differential strengths of an *x* solution habit and negative transfer, such an explanation would be unable to account for the fact that the groups which solved $z_1$ most quickly were also the ones which first returned to an *x* solution.

In devising this study no predictions were made regarding differences among the groups at Trial $z_6$, the first trial which was solvable by either *z* or *x* solutions. The differences which were obtained there were somewhat of a surprise although it should be pointed out that the two largest differences are only at the 5% and 8% level of significance and an over-all analysis yields an *F* of only 2.047, significant between the 8% and 10% levels of confidence. The marked slowness of Group 3 is difficult to interpret. Perhaps this group was able to recognize that more than one solution was possible and was disturbed by the conflict of deciding which was correct. This would affect Groups 1 and 2 less since they were less likely to perceive more than one solution as possible, and to disturb Group 4 less since it was more prepared for any kind of change in the situation. Such explanations, however, are obviously post hoc.

Although the training series in this experiment differs from the traditional partial reinforcement design, certain similarities exist. For example, Group 1 takes longer to change its solution on Trial $z_1$, but quickly thereafter accepts a new solution, as indicated by the results on Trial $z_2$. This suggests that the same phenomena may be involved as in the rapid extinction following 100% reinforcement. In a similar fashion there are analogies between Groups 3 and 4 and the results from less than 100% reinforcement. The evidence for partial reinforcement in terms of the slower extinction in the partially reinforced Groups 3 and 4 is also demonstrated by the fact that these groups return more easily to the original *x* solution.

It would seem that the occurrence of a reward following a nonreward for a particular solution not only strengthens or weakens a particular habit, but also increases the expectancy that rewards follow nonrewards. From the point of view taken in this paper, the two situations are similar in that in each, another expectancy is being learned other than the one that leads directly to the reinforcement.

*Application to personality study.*—The description of the behavior of our rigid group fits fairly well the description of rigid behavior in feebleminded subjects given by Kounin (3) as well as other workers. However, we cannot at this time generalize from these findings to clinical groups or to the behavior of maladjusted individuals. We do not know to what extent such behavior is specific to given situations or whether or not the behavior of individuals operating in much more complex social situations can likewise be predicted using the same principles as were used in making predictions for our experimental groups. Nevertheless such principles appear to provide a basis for explaining a similarity of behavior in widely disparate groups such as those considered maladjusted, feebleminded, brain-injured, and institutionalized. In addition these principles have the advantage of going beyond immediate description and defining the conditions which allow for prediction and perhaps therapy. Somewhat oversimplified examples of the application of such principles are given below.

In the case of the feebleminded *S* he is both limited in learning alternative solutions and has frequent experiences of failure which lead him to seize upon any positively reinforced or correct solution and maintain it. The single solution is maintained in spite of its failure in changing situations because of his low expectancy for positive reinforcement from any other pathway. The brain-injured, likewise, operate under physical handicaps which limit the kinds of solutions they can learn and have frequent experiences of failure and inadequacy. In the case of the so-called neurotic or maladjusted individual whose rigidity consists mainly of maintaining some defense or avoidance behavior in the absence of any "real threat," the avoidance behavior itself may be regarded as a single solution which is regularly reinforced by preventing the reoccurrence of some trauma, thus providing the conditions making for single-solution learning. Individuals who spend much time in institutions where there are relatively inflexible rules, precedents, schedules, etc., to determine behavior are in a situation that may be characterized as single-pathway learning.

Some mention should be made here of an obvious application to the problems of therapy. In the case of persistent maladjustment it would seem important that *S* does not merely learn another single solution as better but that he learn and have reinforced a behavior of looking for or seeking alternative solutions. Morton (6) in using a very brief therapy technique demonstrated the effectiveness of such techniques in one of the rare psychotherapy studies in which matched controls were used.

For the time being, however, such speculations regarding application of these results must be considered as hypotheses for further experimental investigation.

## Summary

Adopting a social-learning framework developed by Rotter and his students, the problem of rigidity was approached in this study not as a trait or entity but as a kind of behavior predictable from specific learning experiences. Briefly, it was hypothesized that nonrigidity or flexibility was a kind of higher-level behavior which consists of expecting change and looking for alternative pathways. Rigid behavior is typified by the absence of such learning, or by the expectancy of a single correct solution which remains the same. Four groups, each of 26 *S*s, were tested using a problem-sorting task. The procedure for each group was varied to control all relevant conditions other than differences in training for an expectancy of change and alternative solution. Four different tests of the major hypotheses were supported at significant levels. The groups showed consistent progressive differences in predicted directions. Some possible applications of these results to other similar experimental and social-learning situations were considered.

(Received January 17, 1952)

## References

1. Goldstein, K. Concerning rigidity. *Character & Pers.*, 1943, **11**, 209–226.
2. Harlow, H. F. The formation of learning sets. *Psychol. Rev.*, 1949, **56**, 51–65.

3. KOUNIN, J. J. Experimental studies of rigidity: I. The measurement of rigidity in normal and feebleminded persons. II. The explanatory power of the concept of rigidity as applied to feeblemindedness. *Character & Pers.*, 1941, **9**, 251–282.
4. LEWIN, K. *A dynamic theory of personality.* New York: McGraw-Hill, 1933.
5. LUCHINS, A. S. Rigidity and ethnocentrism: a critique. *J. Pers.*, 1949, **17**, 449–466.
6. MORTON, R. B. A controlled experiment in psychotherapy based on Rotter's social learning theory of personality. Unpublished Ph.D. dissertation, Ohio State Univer., 1949.
7. REED, S. Development of non-continuity learning from continuity learning in the albino rat. Unpublished Ph.D. dissertation, Ohio State Univer., 1949.
8. ROKEACH, M. Generalized mental rigidity as a factor in ethnocentrism. *J. abnorm. soc. Psychol.*, 1948, **43**, 259–278.
9. ROTTER, J. B. Tentative formulation of basic principles for a social learning theory of personality. Mimeo., Ohio State Univer., 1948.
10. ROTTER, J. B. Tentative formulation of some social learning constructs. Mimeo. 1–7, Ohio State Univer., Jan. 1950–June, 1951.

# ASSIMILATION IN THE IMMEDIATE REPRODUCTION OF VISUALLY PERCEIVED FIGURES[1]

JEROME S. BRUNER, ROBERT D. BUSIEK, AND A. LEIGH MINTURN

*Harvard University*

This paper concerns itself with the phenomenon of perceptual assimilation and the possible mechanisms which underlie assimilation. By assimilation we mean the tendency for objects in the environment to appear as much as possible like "normal" or "modal" objects of the class of objects to which they are seen to belong. Assimilation may take a perceptual form; more usually it has been studied in connection with memory changes. In the former case, one studies the change in the apparent characteristics of an object when it is identified as a certain kind of object: a patch of unstable orange appears yellower when its form is that of a normally yellow object (**1**). Assimilation in memory is best illustrated by the well-known Carmichael, Hogan, and Walter study (**3**). Ambiguous figures were presented and labeled as either one or another kind of object. Reproductions after a period of time showed distortions toward the characteristics of the label used.

It is with the continuity of perceptual and mnemonic assimilation that the present study is concerned. Can one show that the mnemonic assimilative effects reported by Carmichael, Hogan, and Walter are evident when one uses a procedure involving immediate reproduction with no delays for the operation of memory effects?

[1] This research was aided by a grant from the Laboratory of Social Relations at Harvard University.

## Design and Procedure

Our 59 *O*s, undergraduates at Harvard and Radcliffe, were subdivided as indicated below, into four subgroups. They were presented at tachistiscopic exposures four of the original Carmichael, Hogan, and Walter figures: the trowel-pinetree, stirrup-bottle, lima bean-canoe, and hat-beehive (see Fig. 1). The first in each pair will be referred to as the "*a*" labels, the second as the "*b*" labels. These figures were chosen because they came as close as any in the original study to being equally "distortable" towards both labels. The figures were exposed singly on a projection screen 14 ft. removed from a 115-v., 150-w. Visual Education Society projector, Model DD, to which was fixed a high-speed Ilex shutter. The room was illuminated by four 40-w. overhead fluorescent lights. The figures were 3.65 apparent ft.-candles; the background 3.98 apparent ft.-candles.

All *O*s were given the following instructions upon being brought into the room and seated. "This is an experiment in the accuracy of perception. I will flash some figures on the screen and ask you to draw what you see." A sample figure was then shown to familiarize them with where on the screen the figures would be pro-

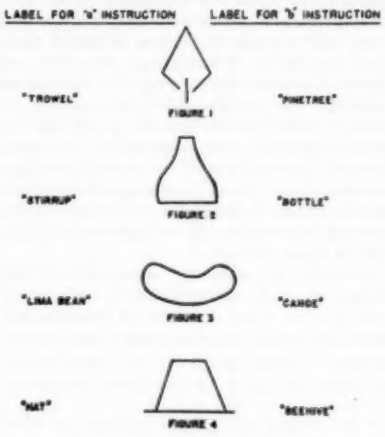


FIG. 1. Stimulus figures

jected, etc. Then, "If you see anything at all, draw it; but if you don't see anything, please let me know and I will re-expose the figure. As you saw, exposures will be rapid, so I'll give you a ready signal before each flash."

Each group then received different instructions. The first of these, the Uninstructed group, UI ($n = 13$), was given no inkling as to what the figures to be exposed would be. "Would you please draw as carefully as you can what you will see on the screen before you. I shall expose it rather rapidly, and ask you to draw it immediately afterwards." After each exposure, *O* drew what he had seen.

The second and third groups, Single Instructed groups, SIa and SIb, were told what each figure would resemble before it was shown. "I am going to show you a figure resembling a —— which I shall ask you to reproduce. Would you please draw as carefully as you can, etc." Group SIa ($n = 13$) was told that the figures resembled respectively a trowel, a stirrup, a lima bean, or a hat. Group SIb ($n = 12$) was told the figures resembled a pinetree, a bottle, a canoe, or a beehive.

The last group of *O*s, the Double Instructed group, DI ($n = 21$), was told the following: "I am going to show you a figure resembling either a —— or a —— which I shall ask you to reproduce. Would you please draw as carefully as you can, etc." Each *O* then was told that a given figure resembled either a pinetree *or* trowel, bottle *or* stirrup, lima bean *or* canoe, hat *or* beehive. Which label came first was determined by chance.

The order in which the figures were presented was randomized. A given figure was first presented at 10 msec. and reproduced by *O*, then at 50 msec. and reproduced, and finally at 100 msec. and reproduced. Then, following these rapid exposures, *O* was shown the figure ad libitum and asked first to copy it as carefully as possible. Then he was asked, for example, to "draw a trowel which looks as much like the figure in front of you as possible." In short, for each figure we have *O*'s drawings at the three exposure levels, his effort at copying the figure exactly, and his drawing of how the figure might be made to resemble the two alternative labels used in the experiment.

*Methods of analyzing assimilation.*—Two judges, operating independently of each other, were used to assess direction of assimilation.[2] Their task was to judge whether a given reproduction of a figure was in the direction of one of its labels, the other of its labels, or toward neither. The judge used a seven-point scale, with zero at its midpoint, and its extremes labeled +3 and −3 respectively. In judging any given reproduction of a given *O*, the scale was anchored for the judge by placing the full-exposure copy of the figure made by *O* at the zero point, and placing at the extreme points his drawings of what the figure would look like if it resembled the two labels, e.g., a pinetree and a trowel.

The judges did not know in what group *O*s belonged and could not know what label they had been given. The *O*s were ordered at random and all reproductions made at 10-msec. exposure were judged first, then those at 50 msec., and finally those at 100 msec.

The computation of inter-judge reliability was carried out on grouped data where all ratings of distortion toward the "*a*" label from +3 to +1 were combined, and all toward the "*b*" label from −3 to −1. We therefore had for each reproduction the possibility of one of three scores by each judge: "*a*" distortion, "*b*" distortion, and "*O*" indicating no distortion. To assess the reliability of agreement we utilized the traditional matching formula discussed by Wilks (**5**, pp. 203–213) and recently extended by Mosteller, which is designed to test the reliability of the difference between number of agreements obtained by two judges and the number to be expected by chance in random matches as in matching two card decks against each other, two stacks of pennies, etc.[3] The matching formula was applied separately to each of the four groups, for each of the four stimuli. For these 16 subgroups of pairs of judgments, chance levels of agreement were exceeded in 12 cases at less than the .02 confidence level, in the other four cases at a confidence level of less than .15.[4] The average expected coincidence of matches for the subgroupings was approximately 16; the average obtained level of joint "hits" was 24.

The scale ratings of the two judges were pooled and it is in terms of combined scores that

[2] The writers are greatly indebted to Mrs. Patience Zawadsky and Miss Patricia Manning for their help in acting as judges.

[3] The formula, in general terms, can be stated thus: if we have $n$ sets of judgments by two independent judges, the probability of $h$ joint hits or agreements is given by:

$$P(h) = 1/h!\,[1/0! - 1/1! + 1/2! - 1/3! + \ldots \pm 1/(n-h)!].$$

If we want to know the probability of $h$ or more hits, we add up the probabilities of the cases of interest, e.g.,

$$P(h) + P(h+1) + \ldots P(n).$$

[4] For a full and tabulated account of the matching reliability procedure as applied to our data, the reader is referred to R. D. Busiek (**2**).

our results will be stated. After averaging the pooled scale ratings, they were again subdivided into "*a*," "*O*," and "*b*" categories respectively, indicating as before assimilation toward one label, no assimilation, and assimilation toward the other label. All reproductions with a pooled rating of "*O*" were, of course, placed in "*O*" category. Since we did not wish to give too much weight to small distortions, we adopted the following practice. Half of the distortions toward "*a*" that had an average rating of .5 were placed in the "*O*" category, half in the "*a*" category—the assignment to category done randomly. The same was done with small distortions toward the "*b*" label.

## Results

The results of our analysis of assimilation tendencies is contained in Table 1. A glance at this table indicates a systematic tendency toward assimilation—SIa tending toward the "*a*" figures, SIb toward the "*b*." We see too that the picture is somewhat obscured by the fact that some of the figures seemed more easily assimilable in one direction than another.

A first approach to a closer look at the assimilation effect is provided by a comparison of the two SI groups. If the members of Groups SIa and SIb are compared on "*a*" assimilation for each of the four test figures at each of the three exposure levels used (10, 50, and 100 msec.), this allows 12 comparisons. In 11 of these 12 comparisons, Group SIa shows more assimilation toward the "*a*" figures than Group SIb. Using the "sign test" developed by Dixon and Mood (**4**), we find this difference between the two groups reliable at the .01 level of confidence.

If we run the comparison in the opposite direction, inquiring whether Group SIb shows more assimilation toward the "*b*" figures than Group SIa, we find again that in 11 of 12 comparisons, SIb does show more assimilation.

TABLE 1

Per Cent of *O*s Showing Assimilation Toward "*a*" and "*b*" Figures or Zero Assimilation

| Group | *n* | Fig. 1 | | | Fig. 2 | | | Fig. 3 | | | Fig. 4 | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | *a* | *O* | *b* | *a* | *O* | *b* | *a* | *O* | *b* | *a* | *O* | *b* |
| *10-Msec. Exposure* | | | | | | | | | | | | | |
| Uninstructed | 13 | 46 | 39 | 15 | 15 | 54 | 31 | 46 | 23 | 31 | 23 | 46 | 31 |
| Single Instr. "a" | 13 | 62 | 23 | 15 | 69 | 0 | 31 | 62 | 23 | 15 | 46 | 23 | 31 |
| Single Instr. "b" | 12 | 25 | 25 | 50 | 25 | 25 | 50 | 17 | 33 | 50 | 33 | 17 | 50 |
| Double Instr. | 21 | 43 | 24 | 33 | 38 | 14 | 48 | 38 | 19 | 43 | 38 | 24 | 38 |
| *50-Msec. Exposure* | | | | | | | | | | | | | |
| Uninstructed | 13 | 69 | 15 | 16 | 38 | 31 | 31 | 15 | 31 | 54 | 31 | 38 | 31 |
| Single Instr. "a" | 13 | 69 | 15 | 16 | 31 | 15 | 54 | 46 | 15 | 39 | 62 | 23 | 15 |
| Single Instr. "b" | 12 | 50 | 25 | 25 | 33 | 25 | 42 | 33 | 8 | 59 | 50 | 8 | 42 |
| Double Instr. | 21 | 62 | 19 | 19 | 28 | 24 | 48 | 24 | 19 | 57 | 38 | 38 | 24 |
| *100-Msec. Exposure* | | | | | | | | | | | | | |
| Uninstructed | 13 | 69 | 31 | 0 | 23 | 39 | 38 | 15 | 16 | 69 | 23 | 62 | 15 |
| Single Instr. "a" | 13 | 46 | 39 | 15 | 54 | 31 | 15 | 54 | 23 | 23 | 54 | 38 | 8 |
| Single Instr. "b" | 12 | 33 | 34 | 33 | 25 | 33 | 42 | 17 | 25 | 58 | 42 | 50 | 8 |
| Double Instr. | 21 | 57 | 39 | 14 | 24 | 19 | 57 | 53 | 14 | 33 | 53 | 33 | 14 |

One may also compare the two SI groups with the UI and DI groups. A comparison of Group SIa with Group UI shows that in 9 of 12 instances, *Os* in the former group showed more assimilation toward the "*a*" figures. This figure, however, falls short of significance in the sign test, being at the .25 level of confidence. A comparison of Group SIa with Group DI shows that in 11 of 12 comparisons, the former shows more assimilation in the expected direction. This result is at the .01 level of confidence. Comparisons of Group SIb with Groups UI and DI yield a 10 to 2 edge for SIb over UI ($p = .05$), and a 9 to 3 superiority for SIb over DI ($p = .25$) in assimilation toward "*b*" figures.

Another estimate of the assimilation effect can be obtained by grouping the two sets of single instructed *Os* (SIa and SIb) and comparing them with the two groups that did not receive single instruction (UI and DI). There is one problem here, however. For some of the figures, there was a greater tendency for assimilation to move in the "*a*" direction, even without instruction, while for others the contrary tendency was noted. This was intrinsic in the nature of the figures and while it was unfortunate, there was little to be done about it. If we combine the two single instructed groups, half of the combined group (SIa) will be given a positive score for "*a*" assimilations, half (SIb) for "*b*" assimilations. This we have done. Two comparisons are possible. For one comparison, a positive score for Group UI is an "*a*" assimilation, for Group DI, a "*b*" assimilation. In a second comparison this is reversed.

In both instances 11 of the 12 comparisons show a higher frequency of predicted assimilation in the SI groups than in the groups not receiving single instructions. These results are at the .01 level of confidence. Stated in highly summarized terms, we may say that at the 10-msec. exposure, somewhat more than half of the single instructed *Os* showed predicted assimilation, while only a third of the others did. The proportions at the other exposure levels are comparable.

Consider finally the difference between the UI and DI groups. We might expect double instructions to increase assimilation either toward the "*a*" or "*b*" figures more than no instruction. This was the case. In 11 of 12 comparisons, Group DI shows a higher frequency of assimilation toward either "*a*" or "*b*" labels. In short, double instructions, conflicting though they may be, induced more assimilative effect than no instructions.

We come finally to the effect of length of exposure on degree of assimilation. The reader will note from Table 1 and from the statement of results thus far that no striking differences were found. Yet, some are to be noted. Take, for example, the amount of assimilation toward "*a*" figures in Group SIa as compared with all other groups. At 10 msec. it is greater than for all other groups. The same can be said for "*b*" assimilation in Group SIb as compared with all others at this exposure level. In 12 comparisons all are in the predicted direction. This is in contrast with the 50-msec. exposure level. A comparison of "*a*" assimilation in Group SIa with all others shows only 8 in 12 in the predicted direction. Where "*b*" assimilation is concerned, 10 in 12 comparisons of Group SIb with other groups are in the predicted direction. At 100 msec., a comparison of Group SIa with Groups SIb, UI, and DI for "*a*" assimilation shows 10 in 12 in the predicted direction. A similar comparison of Group SIb with other groups for "*b*" assimilation yields 7 in

12 in the predicted direction. There appears, in short, to be a slight trend in the predicted direction, but it is not conspicuous: 24 in 24 predicted at 10 msec., 18 in 24 at 50 msec., and 17 in 24 at 100 msec. Perhaps the range of exposure times used was not great enough or, possibly, time is not a very critical variable in the present kind of situation.

## Summary and Discussion

The assimilative memory changes noted in the Carmichael, Hogan, and Walter experiment (**3**) have been shown to operate not only for reproductive memory but also in the immediate reproduction of tachistoscopically presented figures from the Carmichael series. When an ambiguous figure to be shown to *O* is labeled in advance of its exposure, it tends to be reproduced more like its given label than when it is not labeled prior to exposure. Moreover, the same figure, labeled differently for two groups, will tend to be reproduced differently by the two groups. We would like to propose that the same mechanisms which mediate assimilation in "reproductive memory" over time operate to produce the kinds of changes in "reproduction" of immediately presented figures which have been reported in this paper. The objective of this paper was to demonstrate comparable assimilative changes in a conventional memory experiment and in a perceptual experiment—utilizing identical stimulus materials.

Various precautions were taken to be sure that assimilative changes were in the direction of *O*'s conception of what figures variously labeled would look like. It was not possible, however, to control the effect of prior labeling upon motor reproductive tendencies and we may assume that part of the effect can be attributed to this source.

The effect of instruction or labeling prior to the input of a stimulus-to-be-perceived is to increase the likelihood of the categorization of the percept as having a particular identity. Identifying a figure is a matter of relating it to a category of objects. Any given category of objects may be conceived to have an adaptation level or, in other terms, a "typical instance." Where stimulus input is inadequate—in the present case, made so through brief exposure—the resulting percept will be assimilated toward the typical instance of a category. The same process may operate in memory: assimilation being a systematic change in the "memory image" in the direction of typicality. Any reduction in stimulus determination, through reduced stimulus adequacy or through an alteration of forces of organization in a memory trace, will bring about this effect.

(Manuscript received November 5, 1951)

## References

1. Bruner, J. S., Postman, L., & Rodrigues, J. S. Expectancy and the perception of color. *Amer. J. Psychol.*, 1951, **64**, 216–227.
2. Busiek, R. D. The effect of prior labelling on the visual appearance of ambiguous figures. Honors thesis in the Department of Social Relations, Harvard University, 1951. On file in the Harvard College Library.
3. Carmichael, L., Hogan, H. P., & Walter, A. A. An experimental study of the effect of language on the reproduction of visually perceived forms. *J. exp. Psychol.*, 1932, **15**, 73–86.
4. Dixon, W. J., & Mood, A. M. The statistical sign test. *J. Amer. Statist. Ass.*, 1946, **41**, 557–566.
5. Wilks, S. S. *Mathematical statistics.* Princeton: Princeton Univer. Press, 1943.

# THE EFFECTS OF FAILURE AND RETROACTIVE INHIBITION ON MEDIATED GENERALIZATION [1]

BENNET B. MURDOCK, JR.

*University of Vermont*

The concept of "mediated generalization" is one that has wide explanatory value in psychology today, but also one that has been subjected to relatively little direct experimental verification. It presumably depends upon what Hull (6) calls a "pure stimulus act," and what Miller and Dollard (8) call "response-produced cues." These two terms refer to essentially equivalent processes, and are applied to a response whose sole function is to serve as a stimulus for other responses. When elicited by any given stimulus, it plays a mediating role by evoking a terminal response which has not been previously evoked by the original stimulus. This type of generalization is common in language behavior, with labels mediating a wide variety of responses.

Mediated generalization has been studied in the context of "mediate association." Peters (10) found that associations tended to transfer from previous pairings (for instance, having learned *A*—*B* and *B*—*C*, *C* would then be evoked by *A*); however, the degree of transfer was a function of the nature of the original pairings and the method of testing for mediate association. More recently, mediated generalization has been studied by Cofer and Foley (2), Razran (11), Riess (12), and Birge (1). The latter study is particularly relevant to the present investigation. Working with children, Birge found that a movement response learned to one object was elicited by a second object with the same name. Furthermore, there was less generalization when *Ss* did not repeat the names aloud in the original learning of the movement response and/or in the test for generalization.

The present experiment assumes that, given the proper conditions, mediated generalization can be obtained. Its chief purpose is to extend our knowledge of the process by determining whether and, if so, under what conditions such generalization can be reduced or eliminated. Miller (7) has recently stated that repression should reduce mediated generalization by removing the necessary verbal labels. This specific prediction is but one of several that follow from the conceptualization of the pure stimulus act as an instrumental act, with all its concomitant functional properties. It could also be predicted that, when this mediating response is subject to experimental extinction, any generalization which depends upon it will also be reduced.

Because of the limitations of the laboratory testing situation, there has been no attempt in this experiment to obtain repression as it is usually defined. Instead, by means of a failure situation an effort was made to induce suppression, a mechanism that is considered by Dollard and Miller (3) to be one part of a continuum having repression itself on the extreme end.

[1] The present study is based on a thesis presented to the Faculty of the Graduate School of Yale University in partial fulfillment of the requirements for the degree of Doctor of Philosophy. The writer is deeply indebted to Dr. Leonard W. Doob for continued advice at every stage of the study; he would also like to thank Dr. Irvin L. Child and Dr. Neal E. Miller for their helpful suggestions and criticisms.

To effect experimental extinction, a retroactive inhibition sequence was also introduced, since this process presumably reduces the strength of the connections between the stimuli and the potential mediating responses.

The present experiment, then, is designed to test the hypotheses that (*a*) suppression, and (*b*) experimental extinction, reduced mediated generalization.

## Procedure

The experimental procedure is shown in Table 1 and presented in detail below. In general, the sequence was as follows: *S*s learned the same name for each of two pictures; they learned to press a push button while naming one of the pictures; they succeeded (or failed) in a task requiring the use of these names; they experienced (or did not experience) retroactive inhibition of the names for the second picture of each pair; and finally they were tested with the second picture of each pair to determine whether the correct button response (although never before presented together with that picture) would occur.

*Subjects.*—All *S*s were tested individually. They were 64 Yale undergraduates in their first course in psychology. The *S*s were told simply that the purpose of the study was to investigate rote learning with nonsense material. Any questions other than procedural were deferred until the end of the experimental session.

*Stage 1: Name learning.*—Each of five nonsense syllables was paired with two of ten different nonsense pictures in a paired-associate learning experiment. The ten pictures were from Gibson (4) and were selected on the basis of minimum similarity from 20 pictures used in a preliminary experiment. The nonsense syllables were from those with 27% association value (5). For purposes of control two different matchings of picture-syllable units were made at random and half of the *S*s worked throughout the experiment with each. This control variable is referred to as Picture—Original Name Matching. The picture-syllable matchings were learned by the anticipation method (2 sec. for S and 2 sec. for S—R) by means of a Hull memory drum. Training was to a criterion of two consecutive perfect trials (20 anticipations). Three different orders of presentation of the matchings were used. The order was random, subject to the restriction that the same picture or syllable did not occur twice in a row.

*Stage 2: Button learning.*—One of the pictures of each pair attached to a nonsense syllable (e.g., $S_1, S_2, \ldots, S_5$) was associated with the pressing of one of five push buttons. For control, two different matchings of pictures and buttons were made at random and half the *S*s worked throughout the experiment with each. This control variable is referred to as Picture—Button Matching. The buttons were on a table in front of *S*, and arranged along a straight line perpendicular to the line of sight between *S* and the memory drum. When a picture appeared in the memory drum, *S* called out its name and then pressed buttons until the correct one was pressed, at which time the shutter closed and the picture disappeared. If *S* neglected to call out any name, he was again instructed to do so; if he had forgotten a name, he was told the name after he had pressed the correct button for the picture. The errors were scored by hand. Training continued to a criterion of two consecutive errorless trials.

*Stage 3: Serial practice.*—As preparation for the success or failure experience, this part of the experiment was structured as practice for an "intelligence test," given "in order to correlate ability to learn meaningless material with intelligence." The test involved the serial learning of the five nonsense syllables ($R_1, R_2, \ldots, R_5$) that had previously been learned to the pictures. The syllables were presented on the memory drum at a 2-sec. rate and learned by the anticipation method to a criterion of two successive errorless trials. In order to make this task more realistic as a measure of intelligence, *S*s were told that the arrangement of names in each list formed a pattern. They were to try to find the pattern, and were told that this was essentially a test of abstract conceptual ability.

*Stage 4: Suppression training.*—At this stage *S*s were divided into success and failure groups of equal $n$ (32) and all *S*s learned a listing of 25 nonsense syllables used in Stage 1 ($R_1, R_2, R_3, R_4, R_5$). This list involved repetitions of each of the five nonsense syllables in the following order as determined from a table of random numbers: 2,2,3,5,1,5,1,4,1,1,1,3,5,3,3,1,2,2,4,2,3,1,2,1,4. The *S*s were told that there was a predictable pattern. Each *S* was given five repetitions of this list, which was presented at a 2-sec. rate with no rest between Unit 25 and Unit 1. No *S* gave more than 25 correct anticipations out of a possible 125.

The differential treatments for success and failure were introduced before this list was learned. The success group was told not to expect to attain complete mastery of the list, as it was "very difficult." Its difficulty was stressed at some length. Instead, they were informed that they could learn part of the pattern without learning it all. They were told that they had to achieve a certain criterion, and the test involved

TABLE 1

EXPERIMENTAL PROCEDURE

| Stage 1 | Stage 2 | Stage 3 | Stage 4 | | Stage 5 | | | | Stage 6 |
|---|---|---|---|---|---|---|---|---|---|
| Name Learning | Button Learning | Serial Practice | Suppression Training | | Retroactive Inhibition | | | | Mediation Test |
| Nonsense Pictures → Nonsense Syllables | Nonsense Pictures → Button Pressing | 5-Unit Serial Learning | 25-Unit Serial Learning | | Exper. After | | Control After | | Nonsense Pictures → Button Pressing |
| $n = 64$ | $n = 64$ | $n = 64$ | Success $n = 32$ | Failure $n = 32$ | Success $n = 16$ | Failure $n = 16$ | Success $n = 16$ | Failure $n = 16$ | $n = 4 \times 16$ |
| $S_1, S_2 \rightarrow R_1$ | $S_1 \rightarrow R_a$ | $R_1$ | $R_2$ | $R_2$ | $S_2 \rightarrow R_6$ | | Cartoon | | $S_2 \rightarrow R_a$? |
| | | ↓ | ↓ | ↓ | | | | | |
| $S_3, S_4 \rightarrow R_2$ | $S_3 \rightarrow R_b$ | $R_2$ | $R_2$ | $R_2$ | $S_4 \rightarrow R_7$ | | Reading | | $S_4 \rightarrow R_b$? |
| | | ↓ | ↓ | ↓ | | | | | |
| ... ... | ... ... | $R_3$ | $R_3$ | $R_3$ | $S_6 \rightarrow R_8$ | | | | $S_6 \rightarrow R_c$? |
| | | ↓ | ↓ | ↓ | | | | | |
| ... ... | ... ... | $R_4$ | $R_5$ | $R_5$ | ... ... | | | | $S_8 \rightarrow R_d$? |
| | | ↓ | ↓ | ↓ | | | | | |
| $S_9, S_{10} \rightarrow R_5$ | $S_9 \rightarrow R_e$ | $R_5$ | $R_1$ | $R_1$ | $S_{10} \rightarrow R_{10}$ | | | | $S_{10} \rightarrow R_e$? |
| | | ↓ | | | | | | | |
| | | $R_1$ | etc. | etc. | | | | | |
| Crit. 2 Errorless Trials | Crit. 2 Errorless Trials | Crit. 2 Errorless Trials | 5 Trials | | Crit. 2 Errorless Trials | | | | Crit. 2 Errorless Trials |

measuring the number of trials required to anticipate a certain percentage of the syllables correctly. The criterion was never stated numerically. They were instructed to anticipate as often as possible, even if it was sheer guesswork. At the end, *Ss* in this group were always informed that they had reached the criterion very quickly and praised for their achievement on this intelligence test.

The *Ss* in the failure group were led to believe that they would discover the pattern. They were told that it might appear difficult at first, but that in reality it was an extremely simple pattern which should cause little difficulty. Memorizing was discouraged; the approach suggested was essentially one of insight. At the end of five replications none of the *Ss* showed any real mastery of the list in their anticipations, and with questioning almost all of them admitted spontaneously that they had no idea what the pattern was. The *E* then said that, since the pattern had not been learned with this many trials, additional trials would probably not help much, and with time limitations it was necessary to continue on to the remainder of the experiment. The *E* said nothing voluntarily about the failure on this intelligence test unless questioned. No disparaging remarks were made, nor were the *Ss* told that they had "failed." Almost invariably the *Ss* asked questions about their performance before continuing. The *E* adopted a sympathetic attitude toward the failure, although he tried to introduce casually some comment about the "amazingly" low incidence of failure. Parenthetically, it might be mentioned that it was far more difficult, by and large, to convince *Ss* that they had succeeded. Usually, the success was far less apparent than the failure.

*Stage 5: Retroactive inhibition.*—Immediately after the success-failure situation, half of the *Ss* of the success group and half of the *Ss* of the failure group were given a new learning situation (experimental group). To equalize the time intervals, the remaining halves were given a cartoon book for 4 min. with the request that they select the cartoon most enjoyed (control group). The interpolated learning consisted of the second set of five pictures ($S_2$, $S_4$, $S_6$, $S_8$, $S_{10}$) and five new nonsense syllables ($R_6$, $R_7$, $R_8$, $R_9$, $R_{10}$). For control, two different matchings of picture-syllable units were made at random and half of the *Ss* worked with each. This control variable is referred to as Picture—Interpolated Name Matching. The method of paired associates was again used. The procedure was identical with that of Stage 1. Training continued to a criterion of two successive errorless trials.

*Stage 6: Mediation test.*—The test for mediation consisted of presenting the second set of pictures ($S_2$, $S_4$, $S_6$, $S_8$, $S_{10}$) to all *Ss* after instructing them to learn to press the appropriate buttons. For each picture, the "correct" button was the one previously learned directly in Stage 2 to the other picture with the same name. The procedure was identical with that of Stage 2 except that *Ss* were not instructed to make the nonsense-syllable response. If they asked whether they should use them, they were told to do as they wished. The testing continued to a criterion of two successive errorless replications.

Following the conclusion of the Mediation Test, all *Ss* were asked the following questions: (*a*) "How did you learn the correct buttons in the last step of the experiment?" (*b*) "What were the nonsense syllables that you learned in the experiment?" (*c*) "How would you rate your performance on the intelligence test? Use an 11-point scale with 0–4 representing poor performance, 6–10 good performance, and 5 in between." (*d*) "Did your motivation on the last part of the experiment [Stage 6] change from what it had been when you first learned the buttons [Stage 2]? If no change, 5; if your motivation increased, 6–10; if it decreased, 0–4."

*Design for analysis of variance.*—The experiment involves two conditions of retroactive inhibition (experimental and control) for each of two conditions of suppression (success and failure in serial learning). In addition, there are three stimulus control variables: Picture—Original Name Matching in Stage 1; Picture—Button Matching in Stage 2; and Picture—Interpolated Name Matching in Stage 5 for the experimental group. Thus the basic design is 2×2×2×2×2 with two *Ss* in each cell, except that the two matchings in the interpolated learning are not represented in the control group. The *Ss* were randomly assigned to groups.

## Results

Stage 6 was the test for mediation, when the five pictures of the second set were presented for the first time with instructions to press the buttons. The means and standard deviations of three different measures of mediation are presented in Table 2. In general, the success-control group made more correct responses and fewer errors on the first five pictures of the Mediation Test and required fewer trials to learn to criterion than the other three groups. Only the following differences were significant: (*a*) On the first five

trials the failure-control, success-experimental, and failure-experimental groups made significantly fewer correct responses than the success-control group ($t$'s of 2.91, 3.46, and 3.44 respectively; in all cases $.01 > p$, one tail). (*b*) On the first five trials the same results obtained for errors ($t$'s of 2.60, 3.18, and 3.15 respectively; in all cases $.01 > p$, one tail). (*c*) For number of trials to learn to criterion the results were in the same direction, but the only significant difference was between the success-control and failure-experimental groups ($t = 1.70$; $.05 > p$, one tail).

Of the three possible measures of mediation, that of number of trials to learn to criterion is least satisfactory because mediation is confounded with learning. (Presumably it is for this reason that fewer significant differences were found with this measure.) Of the two measures utilizing only the first five trials (each picture presented once) the error measure is more satisfactory than correct responses because (*a*) it is more sensitive (i.e., a range of 0–10[2] as opposed to 0–5); and (*b*) the theoretical distribution of chance with which the obtained results can be compared follows a normal curve only for errors. Therefore, the error measure was selected for further analysis.

TABLE 2

MEANS AND *SD*'S FOR THREE MEASURES OF MEDIATION ON MEDIATION TEST: (A) CORRECT RESPONSES ON FIRST FIVE TRIALS; (B) ERRORS ON FIRST FIVE TRIALS; (C) NUMBER OF TRIALS TO LEARN TO CRITERION

| Measure | | Retroactive Inhibition | | | |
|---|---|---|---|---|---|
| | | Control | | Experimental | |
| | | Mean | *SD* | Mean | *SD* |
| Success | A | 3.5 | 1.07 | 2.3 | 0.97 |
| | B | 1.9 | 1.79 | 3.9 | 1.82 |
| | C | 16.5 | 6.09 | 18.1 | 4.42 |
| Failure | A | 2.4 | 1.06 | 2.3 | 0.85 |
| | B | 3.6 | 1.87 | 4.0 | 2.00 |
| | C | 17.6 | 4.50 | 19.9 | 4.96 |

TABLE 3

ANALYSIS OF VARIANCE OF ERRORS ON FIRST FIVE TRIALS OF MEDIATION TEST

| Source of Variance | *df* | Variance | *F* |
|---|---|---|---|
| Success-Failure | 1 | 13.1 | 3.65* |
| Retroactive Inhibition | 1 | 23.7 | 6.60* |
| Stimulus Controls: | | | |
| Picture—Original Name Matching | 1 | 0.0 | — |
| Picture—Button Matching | 1 | 2.6 | — |
| Picture—Interpolated Name Matching | 1 | 2.6 | — |
| Success-Failure × Picture—Original Name Matching | 1 | 17.0 | 4.74* |
| Residual Variance | 40 | 3.59 | |

* Significant at the 5% level of confidence.

The results of an analysis of variance of the error score are presented in Table 3. All five of the main variables and the one significant interaction are indicated. Both experimental variables were significant, while none of the three stimulus control variables were significant. Thus, failure and retroactive inhibition decreased mediated generalization, while the particular matchings of pictures, buttons, and nonsense syllables had no significant effect on mediation. The interaction of success-failure with Picture—Original Name Matching can be under-

[2] Four errors were possible for the first picture, three for the second, two for the third, one for the fourth, and none for the fifth. This assumes, of course, that *Ss* selected their responses from the possible alternatives confronting them. That is, it assumes that they did not press a button which had been correct for a previous picture; when this happened, it was not counted as an error. (The results were also analyzed with errors of this type included; although these results are not presented here, the two methods of scoring gave almost identical results.)

stood with reference to the latter variable. In all probability, it represents not only a stimulus control variable but also a motivational variable. There was a significant difference in the original learning of these two matchings (Stage 1); use of words in the Mediation Test (Stage 6) was significantly related to this variable; and changes in motivation as revealed by posttest question (d) were also significantly related to this variable. Therefore, an *ad hoc* explanation of this significant interaction is that the effects of success and failure on mediation were a function of the motivational level involved; however, as this finding is only incidental to the main conclusions, no further analysis is presented.

*Use of names in Mediation Test.*— Retroactive inhibition was introduced to decrease the strength of the original naming response, and thus decrease mediation. That the interpolated learning did weaken the naming response is indicated by the following results: (*a*) In posttest questioning (question b) *Ss* in the experimental group recalled significantly fewer of the original names than did those in the control group ($\chi^2 = 11.534$, $.001 > p$). (*b*) The use of names in the Mediation Test was optional (i.e., unspecified). None of the 32 *Ss* in the experimental group said aloud any of the original names, while 18 of 32 *Ss* in the control group said aloud four or five nonsense syllables.[3] This difference is highly significant ($\chi^2 = 22.337$, $.001 > p$). (*c*) None of the *Ss* in the interpolated-learning group gave any indication in the posttest questioning (question a) that they had in any way used the original nonsense syllables in mediation, nor did any of them recognize the principle of mediation.

The failure situation was introduced to decrease the strength of the naming response through suppression. Since no *S* in the experimental group called out the names, the effectiveness of the failure situation in producing suppression can be evaluated directly only for the control group. Eighteen *Ss* in the control group called out four or five names; 12 were in the success-control and 6 in the failure-control group. Of the 14 *Ss* who used less than three names, four were in the success-control and 10 in the failure-control group. This difference is significant at the 5% level ($\chi^2$ with Yates's correction = 3.175, $.05 > p$, one tail). Thus, failure decreased the spontaneous use of nonsense syllables in the Mediation Test. There was no significant difference between the success and failure groups in the posttest recall of the original nonsense syllables ($\chi^2 = .533$). It would appear, then, that repression was not involved; *Ss* in the failure group could, if asked, recall the names about as well as *Ss* in the success group.

That failure decreased mediation through suppression (as opposed to a generalized disruption resulting from failure) is further suggested by the following results: (*a*) In the Mediation Test, the difference between the success and failure groups in the number of faulty discriminations (i.e., pressing buttons correct for previous pictures; see footnote 2) was not significant ($t = 1.21$, $.30 > p > .20$). The number of faulty discriminations, although not counted in the error score, would seem to be a direct measure of generalized disruption. (*b*) An analysis of variance of the interpolated learning shows that there were no appreciable differ-

[3] The distribution of the overt verbalization of the nonsense syllables in the Mediation Test was bimodal. The *Ss* called out four or five syllables (out of five), or two or less; no *S* called out three.

ences between the success-experimental and failure-experimental groups either for the total number of trials required to learn to criterion or for the total number of errors (in both cases, the variance estimate based on success-failure differences was *less* than the residual variance). (*c*) There was a significant correlation ($r = -.53$) between the number of nonsense syllables called out and the number of incorrect button responses for the control group.

As indicated by the answers to post-test question (a) success and failure had no significant effects on the number of *S*s who recognized the principle of mediation ($\chi^2 = .129$). Finally, different instructions for the success and failure groups resulted in different self-attitudes towards performance on the "intelligence test." From post-test question (c), the mean self-rating for the success group was 5.3; for the failure group, 2.1. Numerically, the higher the self-rating the better satisfied *S* was with his performance. An analysis of variance of the self-ratings showed that this difference is highly significant ($F = 43.62$, $.001 > p$).

*Degree of mediation.*[4]—The distribution of errors expected by chance was calculated; the range was from 0–10 with a mean of 5.0. When the obtained results were compared with chance, it was found that the success-control, failure-control, and success-experimental groups did significantly better than chance ($t$'s of 6.60, 2.86, and 2.34 respectively; in all cases $.05 > p$); the failure-experimental group was very close to the 5% level ($t = 1.94$; $.10 > p > .05$). As the success-experimental and failure-experimental groups did not differ appreciably from each other in any way on an empirical basis, they were combined to provide a more reliable measure; it was found that the experimental group as a whole did significantly better than chance ($t = 3.05$; $.01 > p$). As a further check on this finding, an empirical estimate of chance was obtained from an additional 64 *S*s tested in a classroom situation; when compared with this empirical estimate, the experimental group was again significantly better than chance ($t = 2.07$; $.05 > p > .01$).

The above-chance performance of the experimental group indicates that some mediation occurred in spite of the interpolated learning. This result cannot be ascribed to the matchings of the pictures, buttons, and nonsense syllables: none of the stimulus control variables had a significant effect on mediation. Also, a detailed analysis of the responses in the Mediation Test showed that primary stimulus generalization was not involved.

## Discussion

The findings of this experiment substantiate the conclusions of the earlier Birge study (1): mediated generalization can be obtained and measured in an experimental setting. Moreover, this experiment has shown that mediating responses are (at least to some extent) responses that can be modified by learning, and that this modification in turn can affect mediated generalization. In the present study a failure experience reduced mediation by inducing word avoidance or suppression of the potential verbal mediating responses; retroactive inhibition reduced mediation by extinguishing the potential verbal mediating responses.

In order for mediation to occur, there are theoretically two essential requirements: (*a*) The *S*s must use the mediating names or some other common response; (*b*) The *S*s must have retained the correct associations (in

[4] For a more complete analysis, see (9).

this case, the names of the pictures *and* the buttons that are matched with these names). Theoretically, they do not have to be able to verbalize the previous relationships that are employed in the mediation. This assumption is borne out by the results; there were some *Ss* who used the names and pressed the correct buttons almost automatically, and according to the protocols obtained from them afterward were completely unaware of the principle of mediation involved. Furthermore, failure decreased mediation in spite of the fact that it had no effects on the number of *Ss* who recognized the principle of mediation.

It was found that the experimental group showed evidence of mediation even though none of the *Ss* in this group made any overt use of the mediating names. This may perhaps simply indicate that the mediating responses were present but not directly observed. However, an alternative hypothesis is also tenable: the correct responses on the Mediation Test may have been learned *directly* in an earlier stage of the experiment. For instance, the *Ss* in Stage 1 may have rehearsed by visualizing a picture when they thought of the name; the names in Stage 2 which were evoked by the first set of pictures may have evoked the corresponding images of the pictures of the second set; consequently, the button responses could have been learned directly not only to the first set of pictures but also to images of the second set. The design of this experiment precludes an evaluation of the relative merits of these two explanations, but the results of the Peters study (**10**) on mediate association suggest that direct learning may be involved. However, even if direct learning did occur, the results show that the nonsense syllables still played a role in mediating the button responses.


## Summary

1. The present experiment was designed to test the hypotheses that (*a*) suppression and (*b*) experimental extinction reduce mediated generalization. Sixty-four college undergraduates were tested individually in a verbal-learning type of experiment; suppression was attempted by a failure situation, and extinction by a retroactive-inhibition situation.

2. The *Ss* first learned the names (nonsense syllables) of ten (nonsense) pictures; as there were only five names, one syllable was the name of two pictures. Secondly, the *Ss* learned to press a particular push button for each of five pictures. The five pictures were selected so that there was no duplication of names; one picture from each pair was used.

3. The experimental variables were next introduced; with a $2 \times 2$ factorial design, each *S* had one level of each of the experimental variables (success or failure, experimental or control). Failure occurred on a task involving only the five names. The retroactive inhibition involved learning new names to the five pictures not used in the second step.

4. In the test for mediation, the five pictures not used in the second step were presented, and the *Ss* had to learn the buttons to them. Mediation could result in an errorless performance; mediation would occur if the pictures evoked the names and the names evoked the responses (i.e., if the names mediated the responses).

5. It was found that both failure and retroactive inhibition reduced mediation. Furthermore, there is direct evidence that failure did evoke suppression and that retroactive inhibition did produce extinction.

6. Even though the experimental

group made no use of the mediating responses (names), their performance was still significantly better than chance.

(Received December 13, 1951)

## References


1. Birge, J. S. Verbal responses in transfer. Unpublished Ph.D. dissertation, Yale Univer., 1941.
2. Cofer, C. N., & Foley, J. P. Mediated generalization and the interpretation of verbal behavior: I. Prolegomena. *Psychol. Rev.*, 1942, **49**, 513–540.
3. Dollard, J., & Miller, N. E. *Personality and psychotherapy*. New York: McGraw-Hill, 1950.
4. Gibson, E. J. Retroactive inhibition as a function of degree of generalization between tasks. *J. exp. Psychol.*, 1941, **28**, 93–115.
5. Glaze, J. A. The association value of nonsense syllables. *J. genet. Psychol.*, 1928, **35**, 255–267.
6. Hull, C. L. Knowledge and purpose as habit mechanisms. *Psychol. Rev.*, 1930, **37**, 511–525.
7. Miller, N. E. Theory and experiment relating psychoanalytic displacement to stimulus response generalization. *J. abnorm. soc. Psychol.*, 1948, **43**, 155–178.
8. Miller, N. E., & Dollard, J. *Social learning and imitation*. New Haven: Yale Univer. Press, 1941.
9. Murdock, B. B., Jr. The effects of failure and retroactive inhibition on mediated generalization. Unpublished Ph.D. dissertation, Yale Univer., 1951.
10. Peters, H. N. Mediate association. *J. exp. Psychol.*, 1935, **18**, 20–48.
11. Razran, G. H. S. Semantic, syntactic, and phonetographic generalization of verbal conditioning. *Psychol. Bull.*, 1939, **36**, 578. (Abstract)
12. Riess, B. F. Semantic conditioning involving the galvanic skin reflex. *J. exp. Psychol.*, 1940, **26**, 238–240.

# RETROACTIVE INHIBITION IN RECALL AND RECOGNITION[1]

LEO POSTMAN

*University of California*

Amount of retroactive inhibition (RI) varies with the method of measuring retention. The method of recognition has generally yielded smaller decrements in retention than other methods such as recall and relearning (**1**, p. 421 f.). The fact that at least some RI can be produced by the recognition method (**2, 6, 8**) suggests, however, that recall and recognition are not fundamentally distinct processes and are subject to similar forms of interference.

To account for the differences in the amount of RI observed, we must identify those characteristics of recognition and active reproduction which make them differentially susceptible to interference. Analysis of these retention situations suggests that recognition tests as commonly used are *less conducive to competiton between responses* based on original and interpolated learning than are other tests. Since such competition is generally considered a major factor in RI (**4**, pp. 488–495; **7**), a method which minimizes it should be expected to produce relatively small decrements in retention.

When retention of the original activity is measured by the recognition method, competition of responses, in the sense in which it occurs in reproductive memory (e.g., in the methods of paired associates and retained members) cannot occur. Only two general types of responses—recognition and nonrecognition—are required in the retention test. There may, however, be another type of response competition, *viz.*, the arousal of competing recognition responses by original and interpolated stimuli. Performance in the recognition test depends on *S*'s ability to discriminate between originally learned and incorrect items. If the incorrect items were drawn from the interpolated series, a competition between recognition responses could result which would be analogous to the competition between responses in situations requiring active reproduction. Typically, however, the incorrect choices in the recognition test are not taken from the interpolated material but are entirely new items to which *S* has not been previously exposed. Hence, such tests have not been conducive to the arousal of competing recognition responses. The small amount of RI obtained in recognition experiments may be largely due to the absence of such competition.

It follows from this analysis that the small amount of RI in recognition is not due to intrinsic characteristics of recognition memory but has resulted from the nature of the specific tests which have been used. We should expect, therefore, the amount of RI to increase and to approach that obtained by methods of active reproduction if interpolated items are deliberately pitted against original items in the recognition test. The experiment to be reported in this paper tests this hypothesis.

## Procedure

*Experimental design.*—The conventional design for RI experiments was employed. There were three main groups which differed only with

[1] The writer wishes to express his thanks to Mrs. Laura W. Phillips and Mrs. Barbara Redalia for their assistance.

respect to the nature of the tests used to measure retention of the original activity. With the *Recall Group*, the method of retained members was used. In *Recognition Group I*, retest of the original learning was by a recognition test in which no member of the interpolated series was presented. In *Recognition Group II*, the recognition test required discrimination between items from the original and interpolated series. Each of the three main groups was subdivided into an experimental group and a control group.

*Original learning.*—The original learning series for all groups consisted of 24 nonsense syllables with an average association value of 25% according to Glaze's norms (3). The syllables were printed in solid capitals (2 in. high) on white cards (4 × 7 in.). The cards were exposed to groups of *S*s at a 2-sec. rate. The series was presented four times, each time in a different random order, with an interval of 10 sec. between series. The following instructions were read to the *S*s before the first exposure of the series: "I am going to show you, one by one, a list of 24 nonsense syllables. I want you to look at them carefully and try to remember as many of the syllables as you can. I shall show the list several times, and then your memory for the syllables will be tested. I shall vary the order in which the syllables are shown from trial to trial, so do not try to remember them in any special order."

*Test of original learning.*—Immediately following original learning, members of the Recall Group were given the following instructions: "Write down as many of the syllables as you can remember from the series which I have just shown you. Write them down in any order that you wish. If you are not entirely sure about any one syllable, write it down anyway. Work fast since there is a time limit." The *S*s were allowed a maximum of 5 min. for recall.

Members of Recognition Groups I and II were given a recognition test consisting of 24 multiple-choice items. Each multiple-choice item consisted of four syllables, one of which was taken from the learning series. None of the incorrect items had more than one letter in common with any of the correct syllables. The following instructions appeared at the top of the test sheet: "This is a multiple-choice test. One of the syllables in each set was among those shown to you before. Encircle the syllable which you recognize as having been shown to you. If you are not sure, guess, but be sure to encircle *one* syllable in each set." A time limit of 5 min. was again imposed. The *S*s were instructed to turn their test sheets over as soon as they had finished.

*Interpolated learning.*—The interpolated series learned by all the experimental groups again consisted of 24 nonsense syllables, with an average association value of 25%. None of the interpolated items had more than one letter in common with any one syllable in the original series. The learning procedure was identical in all respects with that used during the original learning.

The control groups rested for a period of time equal to that occupied by the interpolated learning for the experimental groups. During the rest period, the control *S*s read a selection from the *New Yorker* magazine. The *E* informed the *S*s that he was interested in the amount of material they could read at their normal reading speed during the allotted time.

*Retest.*—The retention of the original learning series was retested immediately following the interpolated learning or the interpolated-reading period.

The Recall Group retest procedure was identical with that used in the first test of the original learning, except for the additional instruction to recall the first of the two lists. Similar additional instructions appeared in the recognition tests.

The composition of the retests for Recognition Groups I and II was the same as that of the original test except for two changes: (*a*) a different random order of multiple-choice items was used; (*b*) in each of the multiple-choice items, one of the incorrect syllables was different from those

TABLE 1

MEAN NUMBER OF ITEMS CORRECTLY RETAINED ON ORIGINAL TEST AND RETEST

| Condition | Experimental Group | | | | | Control Group | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | *N* | Original Test | | Retest | | *N* | Original Test | | Retest | |
| | | Mean | *SD* | Mean | *SD* | | Mean | *SD* | Mean | *SD* |
| Recall | 51 | 8.63 | 3.11 | 6.35 | 3.47 | 52 | 8.03 | 3.55 | 7.45 | 3.10 |
| Recognition I | 64 | 21.59 | 2.90 | 19.41 | 3.94 | 71 | 21.52 | 2.68 | 21.03 | 2.95 |
| Recognition II | 87 | 21.55 | 2.80 | 18.57 | 3.86 | 68 | 21.84 | 3.53 | 21.54 | 2.39 |

used in the first test. For Recognition Group I, the new incorrect syllable again did not have more than one letter in common with any original or interpolated syllable. In the case of Recognition Group II, the new incorrect syllable was taken from the interpolated list. For Recognition Group I, therefore, each multiple-choice item called for a discrimination between an original syllable and three incorrect syllables, two of which had appeared in the first test and one of which was new. For Recognition Group II, each multiple-choice item required a discrimination between an original syllable, an interpolated syllable, and two wrong syllables which had been used in the first test. It is this difference in the nature of the retest of the original activity which constitutes the major experimental variable.

*Subjects.*—The *S*s were 393 undergraduate students at the University of California. They did not know the purpose of the experiment. The numbers of *S*s in the different groups are shown in Table 1.

## Results

*Original learning.*—The original-learning scores of the various groups are shown in Table 1. Recognition scores are, as usual, considerably higher than recall scores. The experimental and control groups are in each case well matched in original-learning performance, and in no case does the difference between any two groups approach statistical significance.

*Retention loss after interpolation.*—The retest scores of all groups appear in Table 1. The percentages of RI for each condition, as shown in Table 2, were obtained as follows: For each of the groups, the mean number of items correct on the retest was expressed as a percentage of the mean number correct on the original test. Using these percentages, the conventional formula, 100 (rest-work)/rest, was applied.[2]

Table 2 also shows tests of significance of the differences in retention loss between experimental groups and control groups. The mean decrease in number of items correct from original test to retest was determined for each group. The $t$ values shown in Table 2 are based on the differences between experimental groups and control groups in the mean number of items lost.

TABLE 2

Per Cent RI under Different Conditions and Significance of Differences in Retention Loss between Experimental and Control Groups

| Condition | %RI | $t$ | $p$ |
|---|---|---|---|
| Recall | 20.69 | 4.61 | <.01 |
| Recognition I | 8.00 | 6.07 | <.01 |
| Recognition II | 12.63 | 6.27 | <.01 |

All experimental conditions yielded significant amounts of RI. The retention decrements do, however, vary systematically as a function of the method of measuring retention. The Recall Group shows the highest per cent RI; Recognition Group I, the lowest, with Recognition Group II occupying an intermediate position. The difference in retention loss between the two Recognition Groups is significant ($t = 1.93$, $p = .03$) by the one-tail test appropriate in this study.

*Intrusions.*—For the 51 experimental *S*s in the Recall Group there were 30 intrusions of interpolated items in the recall of the original series, i.e., an average of .59 intrusions per *S*. Two *S*s from the control group gave "intrusions" through chance error. Only those intrusions which were actually syllables from the interpolated list were included in this count. There was, in addition, a considerable number of "partial intrusions," i.e., wrong syllables which resembled members of the interpolated series but in which one of the letters had been changed. Since the classification of such errors as intrusions is uncertain, however,

[2] It seemed preferable to use these percentages rather than raw retest scores in the computation of RI since the original test scores were very different for the recall and recognition groups.

they were not included in the count.

In the case of Recognition Group II, the analogue to overt intrusions at recall is the choice of interpolated rather than original syllables on the recognition test.

Some interpolated syllables may, of course, be chosen by chance, especially since *Ss* were forced to guess. An estimate of the chance expectation of such choices is provided by the frequency with which the experimental *Ss* in Recognition Group I checked the new syllables which corresponded to the interpolated syllables for Recognition Group II. The experimental *Ss* in Recognition Group II checked an average of 2.72 interpolated items or 50.1% of all incorrect choices; those in Recognition Group I checked an average of 1.23 new syllables or 26.8% of all wrong choices. Clearly the experimental *Ss* of Recognition Group II show a systematic tendency to favor interpolated items among their wrong choices.

In the case of Recognition Group I, no analogue to overt intrusions could be measured since the interpolated items were not represented in the recognition test.

## Discussion

*Retroactive inhibition in recall and recognition.*—The results of the experiment confirm the view that recall and recognition do not differ fundamentally in their susceptibility to RI. Differences in amount of inhibition observed may be attributed, to a considerable extent, to the particular characteristics of the testing procedures.

When the method of retained members is used, the test situation is conducive to competition between original and interpolated responses. Overt intrusions provide direct evidence for such competition. An analysis of interlist interference by McGeoch (5) suggests the probable mechanism. Each member of both original and interpolated series is learned in two contexts: (*a*) a common environmental context provided by the various features of the experimental situation, and (*b*) an internal or representative context provided by the associations and interpretations specific to the item. When the internal contexts of both original and interpolated items are re-aroused during the retention test, there may be competition between the two types of responses.

McGeoch's analysis may be extended to the recognition situation. If both original and interpolated items are present in the recognition test, the internal contexts of both may be re-aroused simultaneously and competition between recognition responses may result. The extent to which such competing recognition responses are aroused depends on the number of interpolated items (or items sufficiently similar to them) included in the retest.

The Recall Group shows an even higher amount of RI than Recognition Group II. Even with the competition factor added, recognition remains a somewhat less sensitive index of RI. The simultaneous presence of original and interpolated items not only arouses competing recognition responses; it also provides additional rehearsal for both types of items and thus may restore to some degree the differentiation between the lists which was weakened during interpolation.

*Relation to two-factor theory.*—Melton and Irwin (7) have proposed a two-factor theory according to which the observed amount of RI depends on both (*a*) the unlearning of the original activity during interpolation and (*b*) competition of responses at recall. This theory has found its

clearest application to tests of reproductive memory which provide opportunities for competition between original and interpolated responses. If the term "competition" is interpreted to include competition between recognition responses, our results offer support for a two-factor theory. For Recognition Group I, there was little possibility of response competition, yet there was a substantial and significant amount of RI. This interference may be ascribed largely to the unlearning of the original activity during the interpolated learning. In the case of Recognition Group II, competition between recognition responses is superimposed on the effects of unlearning and leads to higher RI. The difference between the two Recognition Groups may, indeed, be considered as a rough measure of the effectiveness of the factor of competition in recognition memory.

## Summary

Recognition tests have typically yielded smaller amounts of retroactive inhibition than tests of reproductive memory. This experiment was concerned with a test of the hypothesis this this difference depends on the differential opportunity provided by recognition and recall for competition between responses based on original and interpolated learning.

Original and interpolated learning consisted of lists of 24 nonsense syllables. With the Recall Group, the method of retained members was used to test retention following interpolation. For Recognition Group I, a multiple-choice recognition test was used in which no member of the interpolated series was represented. For Recognition Group II, a multiple-choice recognition test was used in which each correct recognition required discrimination between an original and interpolated syllable.

The Recall Group showed the highest amount of RI; Recognition Group I, the lowest. Recognition Group II showed significantly more RI than did Recognition Group I. The *Ss* in Recognition Group II showed a systematic tendency to choose interpolated items on the recognition test.

These results confirm the view that recall and recognition are not fundamentally distinct processes and are subject to similar forms of interference. For both methods of measuring retention, the results are in agreement with a two-factor theory ascribing RI to both unlearning and response competition.

(Received December 31, 1951)

## References

1. Britt, S. H. Retroactive inhibition: a review of the literature. *Psychol. Bull.*, 1935, **32**, 381–440.
2. Gibson, J. J. Retroaction and the method of recognition. *J. gen. Psychol.*, 1934, **10**, 234–236.
3. Glaze, J. A. The association value of nonsense syllables. *J. genet. Psychol.*, 1928, **35**, 255–267.
4. McGeoch, J. A. *The psychology of human learning*. New York: Longmans, Green, 1942.
5. McGeoch, J. A., & McGeoch, G. O. Studies in retroactive inhibition: VI. The influence of the relative serial positions of interpolated synonyms. *J. exp. Psychol.*, 1937, **21**, 320–329.
6. McKinney, F. Retroactive inhibition and recognition memory. *J. exp. Psychol.*, 1935, **18**, 585–598.
7. Melton, A. W., & Irwin, J. McQ. The influence of degree of interpolated learning on retroactive inhibition and the overt transfer of specific responses. *Amer. J. Psychol.*, 1940, **53**, 173–203.
8. Zangwill, O. L. The problem of retroactive inhibition in relation to recognition. *Brit. J. Psychol.*, 1938, **28**, 229–247.

# EXPERIMENTAL EXTINCTION AS A FUNCTION OF THE INTERTRIAL INTERVALS DURING CONDITIONING AND EXTINCTION [1,2]

WARREN H. TEICHNER

*State University of Iowa* [3]

Experimental extinction or the decrement in the strength of a conditioned response with failure of presentation of the unconditioned or reinforcing stimulus has been interpreted in a number of different ways. In general these formulations fall into three classes: (*a*) those that assume that some inhibitory state or process is developed during extinction which is responsible for the weakening of the response (**8**); (*b*) those that deny the necessity of postulating a special inhibitory factor and rely solely on a competing, excitatory process (**2, 3, 4, 17, 18**); and (*c*) two-factor theories that make use of both types of processes (**6**).

The experimental literature relevant to these various theories is meager and, for the most part, equivocal. The inhibition theory of Pavlov (**8**) and the two-factor theory of Hull (**6**) have a certain advantage over the interference theory of Guthrie (**3, 4**) and others (**2, 17, 18**) in that they were specifically formulated with a view of handling the various phenomena in conditioning studies coming under the heading of inhibition, e.g., experimental extinction, disinhibition, spontaneous recovery, inhibition of conditioning, etc. The interference theorists have attempted to account for these same findings without assuming some inhibitory concept such as Pavlov's internal inhibition or Hull's reactive inhibition.

One experimental finding that has been cited frequently as offering critical evidence with respect to the interference theory is that variation of the intertrial interval has opposite effects in acquisition and extinction. This, it is claimed, is contrary to the implications of the interference theory since, according to the proponents of such a view (**2, 3, 4, 17**), extinction does not differ from learning but merely involves *relearning* of a new incompatible response. Interference theory would, therefore, be forced to predict that variation of the intertrial interval would have comparable effects on extinction and acquisition. The implication of the inhibition theories (both internal inhibition and reactive inhibition) for such simple learning situations as classical and instrumental conditioning is that this variable would have opposite effects, relatively short intervals slowing down acquisition and speeding up extinction.

Actually the experimental findings on this variable in conditioning studies are somewhat equivocal, particularly as to its effects on extinction. Whereas studies of acquisition (e.g., **1, 7, 13, 16**) have all been more or less in agreement in showing that a relatively long intertrial interval is more efficacious in establishing conditioning than a relatively short intertrial interval, those of extinction have been highly contradictory. Thus, in the classical CR situation, Hilgard and Marquis (**5**), Pavlov (**8**), Razran (**11**), Reyna (**12**), and Reynolds (**13**) have reported finding shorter intertrial intervals speeded up extinction, whereas Porter (**10**) failed to find any difference among five groups extinguished with different intervals. Again in instrumental learning situations (Skinner Box, runway, etc.) Porter (**9**) failed to find any difference among extinction groups, Rohrer (**14**) showed that massing tended to speed up extinction, while Sheffield (**15**) found just the opposite result.

An important factor shown to be operating in these studies of the effects of the intertrial interval on extinction is the relation of the intervals employed in the original learning to those used in the extinction period. Thus, Reynolds (**13**) found that following distributed conditioning massed extinction was quicker than spaced. But following massed conditioning no consistent differences were found between massed and spaced extinction. Reynolds suggested that Porter's

[1] This article is based on a dissertation submitted to the faculty of the Department of Psychology of the State University of Iowa in partial fulfillment of the requirements for the Ph.D. degree. The writer wishes to express his sincere appreciation to Prof. Kenneth W. Spence for his helpful suggestions and criticisms throughout the course of the investigation.

[2] Experiment I was reported at the 1951 meeting of the Western Psychological Association.

[3] Now at the Psychology Branch, Aero Medical Laboratory, Wright-Patterson AFB.

(9, 10) failure to find a difference among his extinction groups was possibly due to the fact that the original conditioning was relatively massed.

Reyna (12) reported a similar interrelationship between the intervals employed during conditioning and those used in extinction. His *Ss* were all conditioned with a 60-sec. intertrial interval. Comparison of subgroups extinguished with 15-, 30-, 60-, or 120-sec. intervals revealed slower extinction the longer the intertrial interval up to 60 sec. No difference in extinction rate was found between the 60- and 120-sec. groups.

While Sheffield's (15) findings with the instrumental situation were opposite to those obtained from the classical conditioning experiment in that the spaced groups tended to extinguish more rapidly than the massed groups, they again showed the results to be a function of the intertrial interval used during original learning. The difference between the spaced and massed groups was significant in the case of *Ss* that learned under the spaced conditioning, but was not for those whose original learning involved massing of the trials.

It is apparent from the preceding review of the relevant data that further experimentation is needed to clarify the relation of the magnitude of the intertrial interval to the rate of extinction. In particular there is a need to investigate this relationship with respect to the conditions of distribution of trials in the original learning of the response. The present investigation was conducted therefore as an attempt to throw additional light on the nature of the functional relationships holding between the rate of extinction and: (*a*) the time interval between successive trials in the extinction series, and (*b*) the intertrial interval employed in the original learning of the response.

## Method

*Subjects.*—The *Ss* were experimentally naive male hooded rats of 90–110 days old from the colony maintained by the Department of Psychology, State University of Iowa. There were 125 *Ss* in Exp. I and 80 *Ss* in Exp. II.

*Apparatus.*—The apparatus (Fig. 1) was an unpainted pine box, 10 in. in height, 12 in. in length, and 6½ in. in width (inside dimensions), the back end of which had a hinged door and the top of which was constructed of hardware cloth. The inside of the box was lighted by a 7-w. bulb which hung through the hardware cloth. A food chute extended from the front wall. This was a piece of tin fashioned into a chute of square crossection and spread out at one end to form a small tray. It was pivoted on a fulcrum such that when "at rest" it was parallel to the floor, the tray being 5 in. above the floor. When a pressure of 15 gm. was applied to the tray, it was forced into a "down" position describing about a 45° angle with the floor, the tray then being 3½ in. above the floor. An electromagnet on the outside of the box held the tray in this down position when it made contact with the end of the chute. The tray itself extended inside the box, but was shielded by a guillotine door which slid up and out of the box. Raising the guillotine door started an electric clock measuring in 1/100 sec. The clock was stopped when movement of

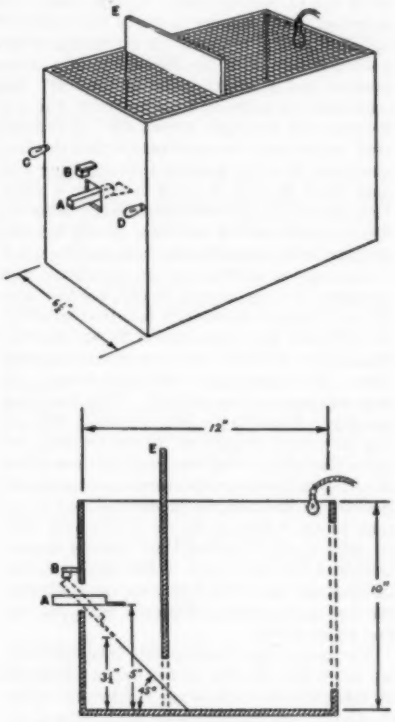


FIG. 1. Apparatus: *A* food chute and tray; *B*, electromagnet; *C*, light source; *D*, photoelectric cell; *E*, guillotine door

the chute to a downward position broke a photoelectric beam. A small pellet of food was inserted into the chute from outside the box. The apparatus was set in a soundproofed room illuminated by a 60-w. bulb in the ceiling approximately 4 ft. from the apparatus.

*Procedure.*—For seven days before the start of the experiment the *S*s were tamed and placed on a 24-hr. feeding schedule consisting of 12 gm. of Purina dog chow given in individual feeding cages. Water was available at all times in both the feeding cages and the home cages. On two separate occasions on the day before beginning training, *S*s were placed singly in the apparatus and allowed to explore for 10 min. One standard-size Purina dog pellet was placed directly before the guillotine door. This door was closed during the whole of this procedure and the 7-w. bulb was not lighted.

On the day following habituation and 23 hr. after their last feeding, *S*s were placed individually in the apparatus and given two preliminary trials and 15 training trials. A preliminary trial consisted of the turning on of the light followed approximately 3 sec. later by the raising of the guillotine door with the food chute in a down position and a .10-gm. pellet in the tray. On removal of the pellet by *S*, the guillotine door was lowered and the light turned off. A training trial was exactly the same except that the food chute was in a rest position thus requiring *S* to push down the tray in order to receive a pellet. Extinction trials followed immediately after the fifteenth training trial and were exactly like the training trials except that no food was presented.

In order to control for the possibility that responses of long latency would interfere with the maintenance of constant intertrial intervals, the guillotine door was lowered 10 sec. after the beginning of a trial if *S* had not responded in this time. If *S* responded in less than 10 sec., the door was lowered immediately. This procedure was begun following the second training trial and was continued throughout all the training and extinction trials. This was justified since it had been determined by preliminary experimentation that the median response latency reduced to less than 10 sec. following the second training trial. In order to allow for the long latencies characteristic of the initial part of the training series, an intertrial interval of 2 min. was used following the first and second preliminary trials and the first training trial.

Extinction began immediately following training and in all cases was carried out to a criterion of three successive failures of response within the 10 sec. period available for the making of the response.

*Experimental design.*—Two experiments were carried out. In Exp. I all *S*s were given 15 training trials with 45 sec. between trials. Subgroups consisting each of 25 *S*s were then subjected to experimental extinction with either 15, 30, 45, 60, or 90 sec. between trials. Assignment to extinction groups was done prior to the experiment by means of a table of random numbers.

A simple factorial design was used for Exp. II. Half of the 80 *S*s used in this experiment were given their 15 training trials with 30 sec. between trials and half with 90 sec. between trials. Immediately following training, *S*s were subjected to experimental extinction in randomly preassigned groups of either 30, 45, 60, or 90 sec. between trials.

## Results

Performance during the learning period was recorded in terms of the latency of response measured from the time of the raising of the guillotine door to the pushing down of the food chute by *S*. The acquisition curves of both experiments in terms of a speed measure (100 × the reciprocal of the median latency for each trial) are presented in Fig. 2. It will be observed that acquisition as measured this way was rapid and approximately linear for the number of trials used. The three curves appear to differ as to slope, the 90-sec. interval curve rising most rapidly, the 45-sec. curve next most rapidly, and the 30-sec. interval curve least rapidly.

A statistical test was performed to ascertain the significance of the differences between the performance of the 30-sec. and the 90-sec. groups. The measure employed was based on the last third of the learning period (Trials 11–15). The median latency for each

TABLE 1

Means and $\sigma_m$'s of the Reciprocals × 100 of the Median Latencies of the Last Third of Learning

| Intertrial Interval | *N* | *Mean* | $\sigma_m$ |
|---|---|---|---|
| 30 | 40 | 35.64 | 1.55 |
| 45 | 125 | 42.34 | 1.34 |
| 90 | 40 | 45.50 | 1.66 |

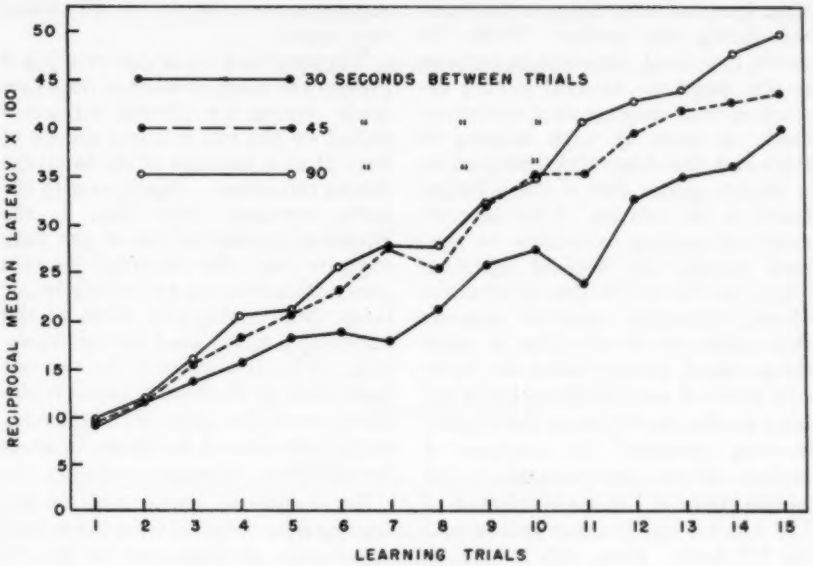


FIG. 2. Strength of response during acquisition

*S* over these last five trials was first determined. The reciprocals of each of these measures were then calculated and the means of their distributions determined. Table 1 shows these means and their standard errors and includes in addition the mean and standard error of the mean of the 45-sec. learning group of Exp. I calculated in the same manner. A *t* performed on the 30- and 90-sec. measures indicated that these intervals produced statistically reliable differences between the two acquisition groups, the null hypothesis being rejected at the 1% level of confidence. This may be taken to indicate that performance during training was a function of the intertrial interval with the level of response strength being lower the shorter the interval.

Curves of performance in terms of latency or the reciprocal of latency could not be determined for the extinction period as it was not possible to compute any measure of central tendency for a group when more than half of the *S*s failed to respond in the 10-sec. trial period. Figure 3 presents a picture of the course of extinction for the five extinction groups of Exp. I plotted in terms of the per cent frequency of responses occurring in the 10-sec. trial period for successive pairs of trials. Inspection of these curves shows that those representing the two shortest interval groups (15 and 30 sec.) are fairly consistently below those for the two longest groups (60 and 90 sec.). The curve for the 45-sec group, which was extinguished at the same interval on which it was originally conditioned, tends to be at about the same level as the two longer intervals.

The middle curve of Fig. 4 indicates how response strength, measured in terms of the mean number of responses occurring in the extinction period for

Exp. I, varied with the intertrial interval during this period. While the curve, in general, rises with an increase in the intertrial interval during extinction thus showing that extinction tends to speed up with massing of trials and slow down with distribution, it would appear that a complicating factor is the relation of the interval employed during extinction to that used during the original learning. Thus, the "bump" in this curve at the 45-sec. extinction interval suggests that resistance to extinction is, other things equal, greater when the intertrial interval used in extinction is the same as that under which the original learning occurred. An analysis of variance of the data presented in the middle curve of Fig. 4 yielded an $F$ of 7.23 which is significant at well beyond the 1% level. From this it may be concluded that the intertrial intervals used during extinction resulted in differences between the mean number of responses occurring in the 20 extinction trials.

The lower and upper curves of Fig. 4 present the mean number of responses made during the 20-trial extinction period by the two learning groups of Exp. II as a function of the intertrial during extinction. Again, as with the curve obtained from Exp. I, the graphical representation of the data suggests that the intertrial interval used during learning was an important factor determining the effect of the intertrial interval used during extinction. This is suggested by the upward turn of the lowest curve at the 30-sec. extinction interval and the relatively high value of the 90-sec. interval for the 90-sec. learning condition.

An analysis of variance of the extinction data obtained from the second experiment as presented in Fig. 4 revealed statistically significant differences only between the learning conditions, $F$ in this instance being 20.07

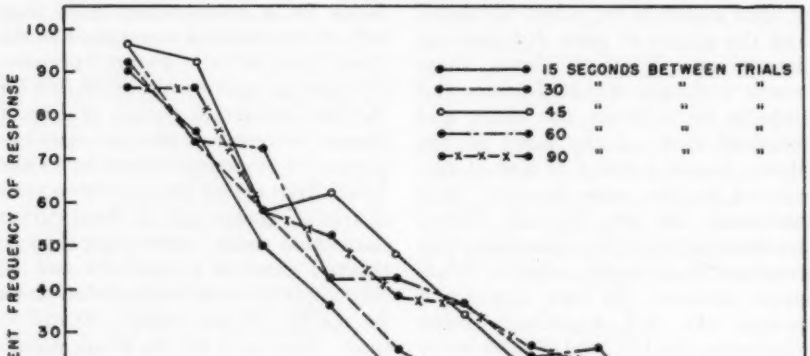


FIG. 3. Per cent frequency of response for pairs of extinction trials under five different intertrial intervals following learning with a 45-sec. interval

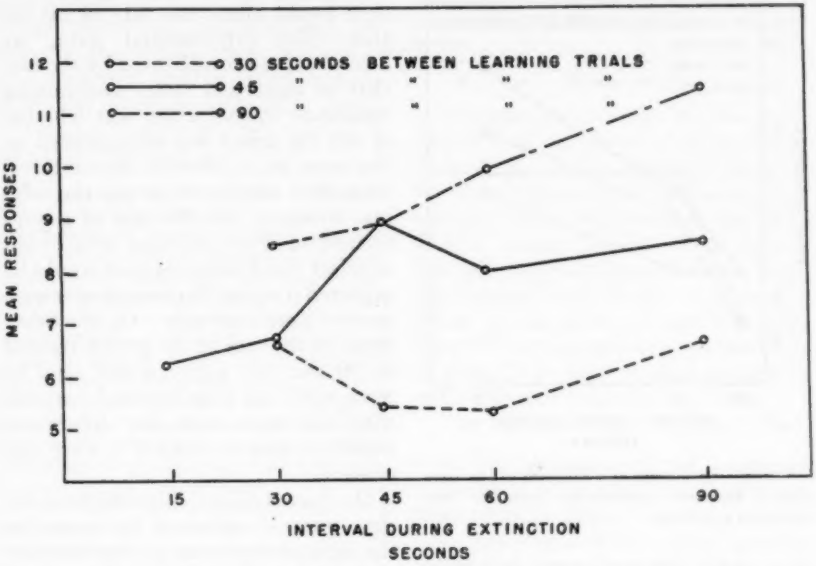


FIG. 4. Extinction as a function of the intertrial interval during extinction for three learning conditions

which with 1 and 72 *df* has a probability of less than .1%. Neither the interaction mean square nor the extinction mean square was significant. However, when a separate analysis of variance was performed on the data obtained from the 90-sec. learning group, it was found that the differences in number of responses among the subgroups cannot be attributed entirely to random sampling within the experiment. The *F* in this case was 7.10 which is well beyond the 1% level of confidence. It will be recalled that a similar analysis of the subgroups trained with a 45-sec. interval obtained from the first experiment also indicated that the intertrial intervals used during extinction had an effect in determining the mean number of responses during the extinction period. An analysis of variance of the extinction groups originally trained with a 30-sec. interval did not yield a significant *F*, indicating that for these groups, at least, the extinction intervals did not result in a difference between the mean number of responses.

It is strikingly clear that the relationship between the rate of extinction and the intertrial interval during extinction is complicated by the relationship between the interval used during extinction and that used during original training. To emphasize this the data of both experiments have been thrown together in Fig. 5 in which is presented the mean number of extinction responses as a function of the intertrial interval during training. Since each of the curves in this figure represents extinction under a constant intertrial interval, the differences reflect the fact that the response was strongest after learning under long intertrial intervals (i.e., 90 sec.) rather

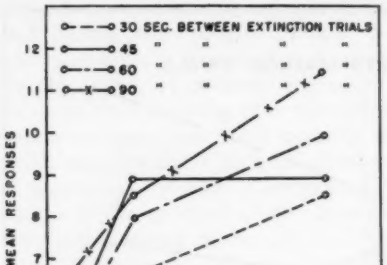


FIG. 5. Extinction as a function of the time interval between conditioning trials for four extinction conditions

than under shorter intervals. It is valuable, moreover, to compare the form of the function of the 60-sec. extinction curve with that of the other three curves since the 60-sec. curve is the only one of the four in which none of the groups represented on the curve experienced the *same* intervals during both learning and extinction. The departures of the other three curves from the form of the curve for the 60-sec. group suggests again that resistance to extinction is greater when the extinction and conditioning intervals are identical. Thus, the 30-sec. extinction curve differs from the general form of the 60-sec. curve in being too high at the 30-sec. value on the abscissa. Similarly, the 45-sec. extinction curve is distorted upward at the 45-sec. interval and the 90-sec. curve at the 90-sec. interval.

## DISCUSSION

Our primary interest was in the question as to whether variation of the intertrial interval during extinction would affect the rate of extinction. The experimental data, we have seen, are complicated by the fact that an important factor determining resistance to extinction was whether or not the group was extinguished on the same or a different time interval from that employed during the original training. In the case of groups trained on 30 sec. this fact would tend to offset the differences that would be expected if extinction were slower with greater time intervals. On the other hand, in the case of the groups trained on 90 sec. this factor would tend to work with the time interval variable with the result that the differences would be greater than if it were not present.

One fact is clear. The effects of the time interval were not the same for the extinction process as they were for acquisition. Whereas acquisition was retarded by the shorter intertrial intervals, the extinction period was either not affected at all or slightly speeded up under the shorter intervals. This finding is not in agreement with those interference theories that assume that the processes involved in acquisition and extinction are identical. On the other hand the data tend to support those theories that postulate an inhibitory state or process since it is possible to deduce from these at least that the effect of the intertrial interval should be different in conditioning and extinction. Thus, following Hull (6), reactive inhibition accumulates with the making of each response and dissipates in the time between responses. The effect of this intervening variable is to depress response strength. With long intertrial intervals greater amounts of inhibition should dissipate and consequently, under such conditions, acquisition should be relatively rapid whereas extinction should be relatively slow.

The present data confirm this expectation clearly for acquisition and, except for the complication introduced by similarities and/or differences between the intervals used during conditioning and extinction, they confirm it for extinction also.

By appealing to the principle of stimulus generalization (6) it is possible to account for the greater resistance to extinction found when the same intertrial interval was used during both acquisition and extinction. According to this principle the effective habit strength of a generalized stimulus is less than that of the CS by an amount determined by the difference in intensity or quality between them. For any trial in a series it is reasonable to suppose that an important part of the stimulus complex consists of (proprioceptive) stimuli persisting from the last previous trial. With a constant intertrial interval this part of the stimulus complex will always be (more or less) the same and it is to these constant characteristics of the stimuli that habit strength accrues. If the intertrial interval is changed, however, from a constant length to some new length, the persisting stimuli and their constant characteristics as well as the stimulus complex will change and accordingly the resultant effective habit strength will be less than the original habit strength. Furthermore, the greater the change in the intertrial interval, the less the resultant effective habit strength and, consequently, the less the strength of response. It follows from this that the greater the *change* (exclusive of length) in the intertrial interval from conditioning to extinction, the more rapid should extinction become. On the other hand, if there were no change in the intertrial interval, the magnitude of effective habit strength would not suffer a decrement, and resistance to extinction would depend then primarily on the amount of reactive inhibition which would dissipate during the trial interval. Since the longer intervals allow a greater dissipation of reactive inhibition, when there is no change in the intertrial interval, the rate of extinction should increase, the shorter the length of the interval.

Although these predictions based on Hull's theoretical formulation are borne out rather nicely by the present data, other conditionable response systems and other conditioning phenomena which also depend on the intertrial interval (e.g., rate of acquisition, reminiscence, spontaneous recovery) must be investigated with similar results before a high degree of confidence can be placed in the generality of the present explanation. Neither of the theoretical constructs employed, habit strength and reactive inhibition, are so defined as to restrict them to the particular conditioning phenomena reported herein.

## Summary

The present investigation was concerned with the course of extinction as a function of the intertrial interval during conditioning and during extinctinction. Two experiments were conducted using hooded rats in an instrumental learning situation. Each experiment involved the training of a group of *S*s with a single intertrial interval and then subjecting subgroups of these *S*s to experimental extinction with different intertrial intervals.

The results obtained indicate that:

1. The acquisition of response strength is more rapid, the longer the time between trials.
2. Resistance to extinction is greater when the same intertrial interval is used during extinction and conditioning than when the intertrial interval during extinction is not the same as the one used during conditioning.
3. Other things equal, the greater the massing of extinction trials, the more rapid the rate of extinction.

The results were employed to evaluate theories of extinction and it was concluded that they do not support interference-type theories. It was found, however, that the general trends of the data could be explained by appeal to Hull's principles of inhibition and the generalization of habit strength.

(Received January 3, 1952)

## References

1. Calvin, J. S. Decremental factors in conditioned-response learning. Unpublished Ph.D. thesis, Yale Univer., 1939.
2. Culler, E. A. Recent advances in some concepts of conditioning. *Psychol. Rev.*, 1941, **29**, 201–216.
3. Guthrie, E. R. *The psychology of learning.* New York: Harper, 1935.

4. Guthrie, E. R. *The psychology of human conflict.* New York: Harper, 1938.
5. Hilgard, E. R., & Marquis, D. G. Acquisition, extinction, and retention of conditioned eyelid responses in dogs. *J. comp. Psychol.*, 1935, **19**, 29–58.
6. Hull, C. L. *Principles of behavior.* New York: D. Appleton-Century, 1943.
7. Humphreys, L. G. Distributed practice in the development of the conditioned eyelid reaction. *J. gen. Psychol.*, 1940, **22**, 379–385.
8. Pavlov, I. P. *Conditioned reflexes.* (Trans. by G. V. Anrep) London: Oxford Univer. Press, 1927.
9. Porter, J. M., Jr. Extinction of an acquired response as a function of the interval between successive non-rewarded trials. *J. comp. Psychol.*, 1938, **26**, 261–270.
10. Porter, J. M., Jr. Experimental extinction as a function of the interval between successive non-reinforced elicitations. *J. gen. Psychol.*, 1939, **20**, 109–134.
11. Razran, G. H. S. The nature of the extinctive process. *Psychol. Rev.*, 1939, **46**, 264–297.
12. Reyna, L. J. Experimental extinction as a function of the interval between extinction trials. Unpublished Ph.D. thesis, State Univer. of Iowa, 1946.
13. Reynolds, B. Extinction of trace conditioned responses as a function of the spacing of trials during the acquisition and extinction series. *J. exp. Psychol.*, 1945, **35**, 81–95.
14. Rohrer, J. H. Experimental extinction as a function of the distribution of extinction trials and response strength. *J. exp. Psychol.*, 1947, **37**, 473–493.
15. Sheffield, V. F. Resistance to extinction as a function of the distribution of extinction trials. *J. exp. Psychol.*, 1950, **40**, 305–313.
16. Spence, K. W., & Norris, E. B. Eyelid conditioning as a function of the intertrial interval. *J. exp. Psychol.*, 1950, **40**, 716–720.
17. Wendt, G. R. An interpretation of inhibition of conditioned reflexes as competition between reaction systems. *Psychol. Rev.*, 1936, **43**, 258–281.
18. Winsor, A. L. Observations on the nature and mechanism of secretory inhibition. *Psychol. Rev.*, 1930, **37**, 399–411.

# SHIFTS IN SCALE VALUES OF ATTITUDE STATEMENTS AS A FUNCTION OF THE COMPOSITION OF THE SCALE [1]

ELIZABETH FEHRER

*Brooklyn College*

If stimuli that can be arranged along a physical or psychological continuum are presented to S in random order, the judgment of each stimulus seems to be dependent in part on the remaining items in the series. Thus the judgment of the heaviness of a weight will depend not only on the absolute magnitude of the weight itself but also on the other weights in the series being judged. If the weight is moved to a different series, the judgment of its heaviness will change. This relativity of judgment has been the subject of numerous studies.[2] Theoretically, the obtained shifts in judgment have been interpreted as a function of shifts in the center of the subjective scale, of extensions of the scale, and more recently in terms of shifts in adaptation level. Of these, the concept of adaptation level seems the most fruitful.

This concept was originally proposed by Helson (**3, 4, 5, 6**) to account for shifts in judgment of sensory stimuli such as colors and weights as the "environment" of these stimuli is altered. Adaptation level (A-L) is defined as the physical value on a stimulus dimension which yields a neutral judgment. In judgments of the saturation of colored samples, for example, the stimulus judged gray would be at adaptation level; in judgments of weight, A-L coincides with that stimulus value judged neither heavy nor light but medium.

Helson has shown that the physical value of the stimulus at adaptation level is a function of such factors as the absolute values of all stimuli being judged, of anchoring (background or comparison) stimuli and of residual (past experience) factors. For example, the A-L for a series of weights is a function of the weighted geometric mean of the stimuli in the series including anchoring stimuli and of residual factors. The psychological or judged value of a given stimulus in a series has been shown to be a function of two factors. It is, first, related to the distance of the stimulus from the A-L value. Second, it depends on the relationship between the physical stimulus values and the psychological judgment values—a relationship which, in accordance with the Fechner law, is logarithmic. From equations embodying this logarithmic relationship and from the knowledge of the location of A-L it is possible to predict the judged value of each stimulus in a series with a very high degree of accuracy (**5**). It is evident that the concept of adaptation level provides a far more precise prediction where it can be applied than explanation by reference to such concepts as the total field, frames of reference, and extensions of subjective judgment scales.

The experiment to be reported below is an attempt to discover whether judgments of such complex social stimuli as

[1] The writer wishes to express her indebtedness to Dr. Harry Helson, to Dr. Evelyn Raskin, and to Dr. James Sakoda for their invaluable suggestions in the carrying out of this study and to Mr. John Grossberg for his assistance in tabulating the data.

[2] Extensive bibliographies may be found in Helson (**4, 5, 8**).

attitude-scale statements will also be influenced by the particular items composing the scale and whether, if such shifts do occur, they can be accounted for on the basis of A-L theory.

Three scales were constructed from Thurstone's scales measuring attitude toward war. The first or control scale contained items evenly spaced over Thurstone's entire range from very pacifistic to very militaristic. The second, the militaristic scale, contained a preponderance of militaristic statements, and the third, the pacifistic scale, a preponderance of pacifistic items. The following hypotheses which derive from A-L theory were to be tested.

1. As A-L is presumably a function of series stimuli, we would expect that A-L (here the items judged neutral) would be different in the three scales.

2. In the scale containing a preponderance of militaristic items, A-L should have a higher value (in terms of degree of militarism) than in the control or pacifistic scales.

3. In the scale containing a preponderance of pacifistic items, A-L should have a lower value (in terms of degree of militarism) than in the control or militaristic scales.

Hypotheses two and three follow from the fact that A-L should be determined by a weighted mean of the series stimuli plus residual factors. If the residual is negligible, then A-L is determined solely by the series stimuli. Residuals will alter this correspondence depending on the value of the residual. Thus, a high residual would result in an A-L above the series mean and vice versa. The residual however should operate as a constant factor in our three scales and therefore if the series is truly effective in determining A-L, A-L should be different in each. In the militaristic scale the mean item is more militaristic than in the control or pacifistic scales. Hence A-L on the militaristic scale should coincide with items that are higher in militarism than the items at A-L on the other two scales.

4. On the militaristic scale, items common to all three scales should be consistently rated more pacifistic than they are on the control or pacifistic scales.

5. On the pacifistic scales, items common to all three scales should be consistently rated more militaristic than they are on the control or militaristic scales.

Hypotheses four and five follow from the fact that the judgment of an item is related to its distance from A-L or the point of neutrality. Thus if the items judged neutral on the militaristic scale are higher in terms of militarism than the items judged neutral on the pacifistic scale, then items common to the two scales should be judged lower in terms of militarism (and therefore more pacifistic) on the militaristic than on the pacifistic scale.

## Method

*Materials.*—The stimuli chosen for study were the statements from the four forms of the Thurstone "Attitude Toward War" scales (9). These are Scale No. 2, Forms A and B, prepared by D. D. Droba and Scale No. 34, Forms A and B, prepared by Ruth C. Peterson. The Thurstone items were particularly desirable for this study as their scale values are already available. (Since these scales were constructed several decades ago it is obvious that the values of the items may be different today. The Thurstone values were therefore used only as an approximation in constructing our scales. The values used in this study were determined directly in this experiment.)

Three scales were constructed from the Thurstone items.

1. A control scale, Scale C, containing five items of each of the Thurstone scale values from zero through ten. The mean Thurstone scale value of these 55 items was 5.45.

2. Experimental Scale M which was weighted towards the militaristic end (Thurstone scale values above five), by omitting the most pacifistic statements. This scale contained 48 items, six each with Thurstone scale values three through ten. The mean Thurstone value of these items was 6.99.

3. Experimental Scale P which was weighted towards the pacifistic end by omitting the most militaristic statements. This scale contained 48 items, six each with Thurstone values from zero through seven. The mean Thurstone value for these items was 3.95.

The following numbers of items were common to the several scales: (*a*) 25 items, five each with values from three through seven, were common to Scales C, M, and P; (*b*) 40 items, five each with values from three through ten, were common to Scales C and M; (*c*) 40 items, five each with values from zero through seven, were common to Scales C and P; (*d*) 30 items, six each with values from three through seven, were common to Scales M and P.

The reasons for constructing scales with such a large amount of overlap were: (*a*) a fairly large number of items common to all three scales would be needed to get a reliable measure of shift, and (*b*) the instructions to *S* to rate the items on a continuum from extreme pacifism to extreme militarism would seem more reasonable when the items in each scale covered a fairly wide range.[3] Because of the large overlap, relatively small shifts in judgment were expected. If, however, reliable shifts were found under these relatively adverse conditions for shifting, the genuineness of the phenomenon would be assured.

The items for each scale were well shuffled and printed on a single sheet with a covering sheet containing instructions. The actual ratings were made on a specially prepared answer sheet. Because of the small number of items, a nine-step rating scale seemed more suitable than Thurstone's traditional eleven categories.

The *S*s were instructed as follows:

"The rating of the statements is to be done on the separate answer sheet. The rating scale has nine steps. Step 1 represents the most pacifistic view possible. If a statement is about as pacifistic as you think a statement could be, place the check mark for that statement under 1. Step 9 represents the most militaristic point of view possible. Those statements that seem to be about as militaristic as they could be should be checked under 9. Step 5 represents neutrality or an opinion that seems neither militaristic nor pacifistic but just about in between. Steps 2, 3, and 4 represent various degrees of pacifism. They are intended to represent equal steps or intervals between neutrality and the most extreme pacifism. In the same way, Steps 6, 7, and 8 should represent equal steps between neutrality and extreme militarism.

"First read over all the statements. Then read them again more carefully, rating them mentally as you go over them. It is not until you have done this that you should record your ratings on the answer sheet. In rating put a check mark in the correct box on the answer sheet after the number of the statement.

"Do not try to put the same number of statements in each rating category. The statements have not been rated as yet. It is on the basis of your ratings that we will know how many belong in each of the steps. Simply put each statement on the step where you feel that it belongs. Use only the steps which express your judgments. All steps need not be used.

"Do not put your name on the answer sheet, but please fill in the blanks under sex and age."

Two points in these instructions should be noted. First, students were directed to read through *all* items before beginning to rate any of them. This was necessary to allow the establishment of an A-L based on all items in the scale before the ratings, which are presumably a function of A-L, were made. Second, it was emphasized that all rating categories need not be used and that the number of items placed in each category need not be the same. It was hoped that this emphasis would avoid any tendencies to judge the items relatively and would instead create a set for "absolute" judgments. The actual results showed that this aim was attained as the frequency with which Categories 1 through 9 were used varied considerably for the three scales.

*Subjects.*—The *S*s were 300 Brooklyn College students enrolled in introductory psychology. Each student was given one scale only, the papers being distributed so that every third student in a row had the same scale. Actually, more than 300 *S*s filled in the scale. The papers used for this study were selected on the following bases: First, papers in which one or more items had not been rated were discarded. Second, papers were so selected that the sex and age composition of the three groups were alike. Each group thus selected contained 100 *S*s, 43 men and 57 women.

*Computation of item scale values.*—Thurstone's method (10) of computing median scale values from plotted frequency curves was not used. Instead median scale values were computed directly from the frequency distributions for each item according to the usual formula, Median $= n/2$. A rating category was taken to represent the interval from its number up to the next higher number. Thus a rating of one represents the interval from one up to but not including two. The point of actual neutrality or A-L would

[3] This is necessary, further, in order not to violate the Michels-Helson "equivalence condition" (8).

have a scale value of 5.5. The psychological distance of an item from A-L is defined by the difference between the scale value of the item and 5.5. For each item the value of $Q$, the interquartile range, was also determined.

## Results and Discussion

### *Computation of Shifts in Adaptation Level*

The data presented in this section will serve to test our first three hypotheses. If A-L differs in our three scales, the items judged neutral on each scale, i.e., those with scale values of 5.5, should be different. Although the extent of shift from one scale to another cannot be measured in physical (cgs) units, it can be measured in psychological units by noting the scale values on a second scale of those items rated 5.5 on a first. For example, if the items rated 5.5 on the M scale are rated 6.5 on Scale C, there has been a downward shift in A-L of one judgment category in proceeding from Scale M to Scale C.

This simple method of measuring shift in A-L can be reliable only if there are many items with scale values close to 5.5. If we define as close to 5.5 items with values varying between 5.0 and 5.9, we find only two on Scale C, the same two on Scale M and these two and one other on Scale P. If we broaden the range of neutral items to include those with values from 4.50 to 6.49, we find nine Scale M items and seven Scale P items falling in this range.

The mean Scale M value of the nine neutral items on this scale is 5.521. On Scale C these same items have a mean scale value of 5.768, indicating that A-L on Scale M is .247 step interval higher than on Scale C. The same items on Scale P have a mean value of 5.738 indicating that A-L on Scale M is .217 higher than on Scale P. These shifts are small, unreliable, and not entirely consistent with our predictions as there should have been a larger shift on Scale P than on Scale C.

The seven items within the neutral range on Scale P have a mean Scale P value of 5.056. Their mean Scale C value is 5.196, indicating that A-L is .14 step lower on Scale C than on Scale P, a finding inconsistent with our third hypothesis. The mean Scale M value of the same items is 4.599, indicating an upward shift of .457, which is in the predicted direction.

The foregoing analysis indicates that A-L does not shift systematically and significantly with a change in the composition of the three scales. The items actually judged neutral are the same on all scales and even the items judged close to neutral are approximately the same. The A-L for stimuli and conditions such as these must be dependent wholly on residual factors.

One of these residual factors may tentatively be identified as consisting of previously learned relations between the items and the judgment categories. As war and peace are issues of intense current concern, our *S*s had had so much previous incidental experience in judging statements similar to those in our scales that large shifts in judgment in our three contexts were improbable. For example, the item, "War is glorious," has so often been identified with strong militarism that its judgment can hardly be influenced by other items in the series of which it is a member.

A second residual factor limiting the amount of shifting may be identified as the semantic relations existing between the items and the judgment categories. Thirteen of the 15 items with scale values of 1 or 9 on Scale C contained specific determiners such as the words "highest," "best," "only," etc. Such extreme words are related semantically to our categories of "extremely militaristic" or "extremely pacifistic," and consequently invite these extreme ratings of the items in which they appear. Thus statements such as "War brings out the *best* qualities in men" or "There is *no* conceivable justification for war" demand extreme judgments which are independent of views expressed in other statements.

The truly neutral items also seem to contain very specific determiners for their placement in this category. The two items with neutral scale values (5.0 to 5.9) on all scales were "War brings out both good and bad qualities in men" and "I never think about war and it doesn't interest me." Both items seem to contain specific determiners. The first item states a clearly intermediate position; the second, although actually seeming an irrelevant item, again does not belong on either side of the neutral point. These items probably defined the neutral point on all scales because of their specific wording and thus prevented a shift in A-L as a function of scale composition.

### *Comparison of Scale Values of Items Common to Several Scales*

Although it has been shown that the same items were rated neutral on all scales and that therefore differences in A-L could not account for differences in the ratings of the other items, still such differences were found in comparing the ratings of items common to two or more scales.

*The 25 items common to all three scales.*—The influence of the series on judgment of individual statements may be seen by comparing the means of the median scale values of the common items when in the three different scales. The Scale C values of these items ranged from 2.750 to 7.594. They do not include items rated very pacifistic or very militaristic by the *S*s who judged Scale C items. The mean scale values of these 25 items, the differences in means and the values of *t* are presented in Table 1. The *t* values were computed by the formula for differences between correlated means of small samples (7, pp. 225–226). This formula is appropriate because we are comparing the effect of two different conditions (i.e., contexts) on the same population (in this case, items).

It may be seen that the shifts, although small, are in the expected direction. Items are rated significantly lower (more pacifistic) when in Scale M than when in Scale C and these ratings in turn are lower (but not significantly lower) than those assigned to the same items when in Scale P. On the average, the 25 items are rated .420 of a category lower on Scale M than on Scale P and this difference is significant beyond the .001 confidence level.

TABLE 1

MEANS AND DIFFERENCES BETWEEN MEAN SCALE VALUES OF THE 25 COMMON ITEMS ON SCALES C, M, AND P

| Scale | Scale Values | | *t* |
|---|---|---|---|
| | Mean | Difference | |
| C | 5.494 | (C−M) .326 | 3.492* |
| M | 5.168 | (C−P) −.094 | 1.043 |
| P | 5.588 | (P−M) .420 | 4.795** |

* Significant between the .01 and .001 confidence levels ($df = 24$).
** Significant beyond the .001 confidence level.

That these 25 items did evidence shifts in judgments may be due to the fact that they are not the ones containing specific determiners for specific judgments. In contrast to the previously discussed Scale C items with values of 1 or 9, almost all of items with Scale C values between 2 and 8 are expressed in far less positive terms and often imply a weighing of issues as, for example, "It is good judgment to sacrifice certain rights in order to prevent war" and "Another war would be terrible but it might be worth the price." Although the wording of these items seems to prohibit the use of extreme ratings and invites instead the use of the intermediate categories, still they do not seem to demand a particular category. They seem sufficiently ambiguous to allow their ratings to vary over a number of categories depending on context. Even in the same context they are rated less consistently as their mean *Q* value is 1.903 in contrast with a mean *Q* value 1.006 for items with Scale C values of 1 or 9. These statements should therefore be the ones to show shifts. Shifts from the militaristic to the pacifistic categories, however, should not be anticipated as the predominance of either pro-war or pro-peace sentiment seems apparent in the specific wording of each statement.

That susceptibility to shift with context should be a function of item ambiguity has been shown

in a number of studies. Asch, Block, and Hertzman, for example, in a study involving a variety of stimuli such as photographs to be ranked for intelligence and honesty and series of occupational titles to be ranked for intelligence and idealism conclude, "We offer the interpretation that when the objective situation is relatively well-structured, or the attitude strongly fixed, that changes of judgment will meet with resistance" (2, p. 250).

The manner in which structuredness is related to shifts in judgment is indicated in another study by Asch (1) in which ten occupational titles including "politics" were to be ranked for several variables. Certain *Ss* were given highly favorable "bogus" ratings for politics; others highly unfavorable ones. The very real suggestion effects obtained were found to be related to the way in which the "bogus" ratings had led the *Ss* to interpret this stimulus item. The *Ss* shown favorable ratings defined politics in terms of Roosevelt, Stimson, etc., while *Ss* shown unfavorable ratings defined it in such terms as Tammany Hall. "According to this view, the characteristic of the process under investigation entails a change in the object of judgment rather than in the judgment of the object" (1, p. 458).

These two studies emphasize that the amount of shift to be expected for an item is related to the characteristics of the item as well as to the characteristics of the series as a whole. It follows that if the items in a series are equal in respect to ambiguity as, for example, in a series of weights, we would expect the judgment of each to shift to an equal extent in a new context. But if the items vary in ambiguity we would expect that even when a general shift in judgment occurs, the extent of shift for each item would be related to its susceptibility to reinterpretation. If, for example, Asch had requested judgments of intelligence for the stimulus "Einstein" it is doubtful whether unfavorable "bogus" ratings of intelligence would have been accepted quite as readily.

As the 25 items common to all three scales exclude those with determiners for extreme judgments and low Scale C ambiguity we cannot relate amount of shift with the presence of these characteristics. Such comparisons, however, were possible with the 40 items common to Scales C and M and with the 40 items common to Scales C and P. In each of the experimental scales the items at the truncated end and in the middle are those with intermediate Scale C values and should shift more than the items at the other end which were assigned extreme judgments in Scale C.

*The 40 items common to Scales C and M.*—These items range in Scale C value from 2.750 to 9.457. They include the extremely militaristic and intermediate statements but exclude the very pacifistic ones.

Table 2 shows the mean scale values of these items when in Scale C and when in Scale M. The difference is small and not significant. A comparison of the mean, however, does not represent the facts adequately as the shifts in judgment varied with Scale C values of the items.

When the items are ranked for Scale C value and divided into four groups of ten each, no differences appear in judgments for the two upper quarters. All of these are items judged militaristic on Scale C. The second quarter, consisting of slightly militaristic, neutral and slightly pacifistic items, shows

TABLE 2

SCALE C AND SCALE M VALUES OF THE 40 COMMON ITEMS

| Comparison | Scale C | Scale M | Difference | |
|---|---|---|---|---|
| | | | (C–M) | *t* |
| Mean scale value (40 items) | 6.746 | 6.564 | .182 | |
| First quarter | 4.050 | 3.596 | .454 | 3.23* |
| Second quarter | 6.105 | 5.819 | .286 | 1.61 |
| Third quarter | 7.665 | 7.670 | −.005 | |
| Fourth quarter | 9.162 | 9.170 | −.008 | |
| Range (40 items) | 2.750–9.457 | 2.759–9.444 | | |

* For *df* of 9, the .01 evel value is 3.250.

a small insignificant difference in the expected direction. The first ten items, all of which are judged mildly pacifistic on Scale C, are judged reliably lower when in Scale M, the difference in average scale value being .454 of a rating category.

These results are not, of course, entirely consistent with the predictions made in Hypotheses 4 and 5, in which it was assumed that the scale values of all items would shift consistently in different contexts. Neither are they consistent with the hypothesis that amount of shift will depend on item ambiguity. Bearing in mind the fact that the fourth or upper quarter includes the statements with low *Q* values and extreme specific determiners, it is not surprising that they are rated the same on both scales. The remaining items, though more ambiguous, do not differ systematically among themselves in ambiguity or *Q* value. The nature of the stimuli cannot, therefore, account for the fact that the shift in the first quarter is greater than that in the second and disappears in the third. It is true that items containing specific extreme determiners and low in ambiguity do not shift and that shift is limited to the ambiguous items. But it is still necessary to explain why certain of these more ambiguous items shift reliably while others do not. An attempt at such explanation will be made following the presentation of the remaining data.

*The 40 items common to Scales C and P.*—The 40 items common to Scales C and P ranged in Scale C value from 1.543 to 7.594. The comparisons involving these items are shown in Table 3. The results are comparable to those found for the 40 items common to Scales C and M but in the reverse direction. Here it is the more militaristic items that are judged differently in the two scales and it is these that are situated at the truncated end of Scale P.

*The 30 items common to Scales M and P.*—The mean Scale M and Scale P values of the 30 items common to these two scales are presented in Table 4. In order to discover which items differ most significantly, the items were first arranged according to the size of their Scale M values and then subdivided into three groups of ten each. Reference to the table shows that the judgments of the items with scale values close to five differ relatively little on the two scales and that the difference is not significant. The more militaristic and pacifistic items, however, do differ significantly. The differences in average scale value for the upper and lower thirds of the

TABLE 3

SCALE C AND SCALE P VALUES OF THE 40 COMMON ITEMS

| Comparison | Scale C | Scale P | Difference | |
|---|---|---|---|---|
| | | | (C–P) | *t* |
| Mean scale value (40 items) | 4.365 | 4.429 | −.064 | |
| First quarter | 2.102 | 2.175 | −.073 | |
| Second quarter | 3.346 | 3.262 | .084 | |
| Third quarter | 5.157 | 4.973 | .184 | |
| Fourth quarter | 6.854 | 7.306 | −.452 | 3.512* |
| Range (40 items) | 1.543–7.594 | 1.562–8.308 | | |

* Significant at .01 confidence level ($df = 9$).

TABLE 4

SCALE M AND SCALE P VALUES OF THE 30 COMMON ITEMS

| Comparison | Scale M | Scale P | Difference | |
|---|---|---|---|---|
| | | | (P–M) | $t$ |
| Mean scale value (30 items) | 5.317 | 5.740 | .423 | 5.392** |
| Lowest third | 3.384 | 3.919 | .535 | 3.149* |
| Middle third | 5.543 | 5.744 | .201 | 2.166 |
| Upper third | 7.025 | 7.559 | .534 | 4.540*** |
| Range (30 items) | 2.759–7.550 | 2.806–8.308 | | |

* Significant between .02 and .01 confidence levels ($df = 9$).
** Significant beyond .001 confidence level ($df = 29$).
*** Significant beyond .001 confidence level ($df = 9$).

items is slightly over one-half of a rating category. It is again evident that the items that shift are those situated at the truncated ends of one or the other of the experimental scales. When all 30 items are considered together, they are judged significantly more pacifistic on Scale M than on Scale P.

The data presented in this section show that significant shifts in scale values occur only for statements located at the truncated end of a scale, i.e., for those intermediate statements in a scale containing no items with extreme specific determiners. The extreme and the truly neutral statements show no shift. The most plausible interpretation of these results seems to be the following one.

First, in line with A-L theory, items are judged according to the extent of their departure from the neutral point. If A-L on different scales is the same because of strong residual factors, then the items judged above this point on one scale will also be judged above it on the other scales. This statement is based on the additional assumption that the items will be placed in consistent rank order regardless of the range of items being ranked. It then follows that if the items with scale values of 5.5 are the same on all scales (whether or not they represent the median of the series) all items ranked above these items will be rated higher than 5.5.

Stimuli, however, are not judged only in accordance with how far they seem to depart from the neutral point. If the values of certain other stimuli in the series are known, the remaining statements cannot take on these values but will be judged in reference to them as well as in reference to the neutral statements. The extremely militaristic and pacifistic items which are semantically tied to the extreme judgment categories probably constitute such additional reference items. When they are present, the intermediate items cannot take on the extreme values but when they are absent, thus leaving the meaning of the extreme categories less well defined, these values may be assigned to some of the intermediate items. As a result, the mean rating of these intermediate items should be more extreme (farther from neutral) in the truncated scale. This reasoning is clearly supported by the data presented in this section.

On Scale M, for example, there were no very pacifistic items that clearly defined the rating Category 1 and that would therefore prevent the use of this category for less extreme pacifistic statements. Examination of Scale M data showed that some *S*s used Category 1 for some of the stimuli so that presumably these stimuli then defined Category 1 for them. (However, *S*s rating the Scale M items used this category only 5% of the time whereas *S*s rating Scale C items used it 14% of the time.) Other *S*s never used this category. For them it obviously represented concepts more pacifistic than those in any of the statements they were asked to rate.[4] As the meaning of Category 1 differed among the *S*s rating the Scale M items, it follows that their ratings of the moderately pacifistic statements should be more variable than the ratings of these statements by *S*s judging them in Scale C.

An examination of the $Q$ values of the items common to Scales C and M supports the hypothesis that unambiguous reference stimuli serve to

[4] The differences among the *S*s' definitions of Category 1 may have arisen from differences in interpretation of the instructions. Even though these emphasized rating rather than ranking, some *S*s may have adopted a ranking rather than a rating set.

define the values of the neighboring stimuli in a series. Table 5 shows that the $Q$ values of the moderately pacifistic items are significantly greater in Scale M than in Scale C. The remaining items common to the two scales, the moderately militaristic (Scale C values 6 and 7) and the extremely militaristic (Scale C values 8 and 9) items do not differ reliably in $Q$ value. A comparable analysis of the $Q$ values of the statements common to the C and P scales yields the same results (Table 6).

Unlike the pacifistic items, the militaristic items in the M scale were rated just as they were in Scale C. In both scales there were strongly militaristic statements to define the rating Category 9 and also the same neutral statements to define the value of 5. In both scales, therefore, these reference items pushed the intermediate statements away from these two scale values and so the intermediate statements were assigned comparable intermediate ratings on both scales.

The chief difference between the present formulation and that of Helson lies in the more explicit emphasis placed here on the influence that reference items which define the two ends of the judgment scale exert on the ratings of the individual items. Helson has not ignored this fact. It is implicit in his demonstration (8) that the slope of the curve relating judgment values with the log of the stimulus values depends on the extent to which $S$ has used all the judgment categories in rating the series stimuli. Slope, in other words, is dependent on the extent to which the Michels-Helson equivalence condition has been met, namely, whether "The judgment-scale and the stimuli encountered are *equivalent* in the sense that the scale is broad enough to include judgments of all stimuli encountered and yet is so narrow that its extreme values do not fall outside the range of judgments elicited by any of the stimuli" (8, p. 357). Our experiment shows clearly that when the equivalence condition is satisfied, there will be relatively little variability in the ratings that different *S*s assign to the same stimuli in the same scale. When it is met for some *S*s but not for others (as in our Scales M and P), there will be significantly greater variability in the absolute ratings assigned to the same items (although these ratings will most probably show a perfect correlation). What is more, this is the condition that permitted shift in the scale values of our stimuli as their context changed. Shifts in the scale values of our items were not caused by shifts in adaptation level as no such shift could be demonstrated. They were determined instead by differences in the interpretation of the extreme judgment categories among our three groups of *S*s.

TABLE 5

$Q$ Values of the 40 Items Common to Scales C and M Classified According to Scale C Values

| Scale C Values | n | Mean $Q$ | | Difference | |
|---|---|---|---|---|---|
| | | Scale C | Scale M | (M–C) | $t$ |
| Below 5 | 11 | 2.075 | 2.359 | .284 | 2.828* |
| 6 and 7 | 13 | 1.758 | 1.908 | .150 | 1.807 |
| 8 and 9 | 14 | 1.283 | 1.264 | −.019 | |

* Significant between .02 and .01 confidence levels ($df = 10$).

TABLE 6

$Q$ Values of the 40 Items Common to Scales C and P Classified According to Scale C Values

| Scale C Values | n | Mean $Q$ | | Difference | |
|---|---|---|---|---|---|
| | | Scale C | Scale P | (P–C) | $t$ |
| Above 5 | 12 | 1.748 | 1.994 | .246 | 2.821* |
| 3 and 4 | 13 | 2.183 | 2.209 | .026 | |
| 1 and 2 | 13 | 1.478 | 1.572 | .094 | |

* Significant between .02 and .01 confidence levels ($df = 11$).

## Summary

The purpose of the present experiment was to determine whether ratings of such social stimuli as attitude statements change as a function of changes in the composition of a scale and whether adaptation-level theory can be applied to the interpretation of such changes.

Three scales were constructed from the items of Thurstone's scales on "Attitude Toward War." Scale C (control) contained an equal number of items of each of Thurstone's values from zero through ten. One of the two experimental scales, Scale M (militaristic), contained items with values from three through ten; the other, Scale P (pacifistic), items with values

from zero through seven. Each scale was rated by a different group of 100 *Ss*, equated for age and sex.

The median scale values of the items common to the several scales differed significantly with context. In Scale M the common items were rated on the average more pacifistic than these same items were rated on Scales C or P. Conversely, in Scale P, these items received the most militaristic ratings.

It was shown, however, that shifts in scale value were limited to certain items. Items rated very militaristic, very pacifistic, and neutral on Scale C maintained the same values on Scales M and P. The items whose scale values changed were those near the truncated ends of the experimental scales, namely, the moderately pacifistic items on Scale M and the moderately militaristic items on Scale P.

As the neutral items were the same on all three scales, the shifts found cannot be attributed to shifts in adaptation level. Instead, they seem to be related to the fact that the extreme judgment categories were differently defined by the three groups of judges.

(Received January 11, 1952)

## References


1. Asch, S. E. Studies in the principles of judgments and attitudes: II. Determination of judgments by group and ego standards. *J. soc. Psychol.*, 1940, **12**, 433–465.
2. Asch, S. E., Block, H., & Hertzman, M. Studies in the principles of judgments and attitudes: I. Two basic principles of judgment. *J. Psychol.*, 1938, **5**, 219–251.
3. Helson, H. Fundamental problems in color vision: I. The principles governing changes in hue, saturation and lightness of non-selective samples in chromatic illumination. *J. exp. Psychol.*, 1933, **23**, 439–476.
4. Helson, H. Adaptation-level as frame of reference for prediction of psychophysical data. *Amer. J. Psychol.*, 1947, **60**, 1–29.
5. Helson, H. Adaptation-level as a basis for a quantitative theory of frames of reference. *Psychol. Rev.*, 1948, **55**, 297–313.
6. Helson, H., & Jeffers, V. B. Fundamental problems in color vision. II. Hue, lightness and saturation of selective samples in chromatic illumination. *J. exp. Psychol.*, 1940, **26**, 1–27.
7. McNemar, Q. *Psychological statistics*. New York: Wiley, 1949.
8. Michels, W. C., & Helson, H. A reformulation of the Fechner law in terms of adaptation-level applied to rating scale data. *Amer. J. Psychol.*, 1949, **62**, 355–368.
9. Thurstone, L. L. (Ed.) *Attitude toward war*, Scale No. 2, Forms A and B (1930), and Scale No. 34, Forms A and B (1931). Chicago: Univer. of Chicago Press, 1930–1931.
10. Thurstone, L. L., & Chave, E. J. *The measurement of attitude*. Chicago: Univer. of Chicago Press, 1929.

# THE EFFECT OF LEARNING ON THE PREDICTABILITY OF PSYCHOMOTOR PERFORMANCE

BRADLEY REYNOLDS

*Perceptual and Motor Skills Research Laboratory, Human Resources Research Center*

It was pointed out in an earlier paper (**10**) that with increased training on psychomotor tests two changes occur in the intercorrelations of trial scores. First, the correlation of early trial scores with remote trial scores decreases as practice increases. Second, the correlation between adjacent trial scores increases with practice. The study to be reported here was designed to study these changes further and to determine whether the correlations of a psychomotor test with a set of printed tests also decrease as a function of amount of training on the psychomotor test. The Complex Coordination Test (**8**) was used since it has a very high ceiling and correlates with many printed tests.

## Procedure

*Apparatus.*—Four copies of the Complex Coordination Test, Form E, described by Melton (**8**, pp. 81–99) were used in the study.

*Printed tests.*—The printed tests used were from the Airmen Classification Test Battery (**2**). The designation and, where necessary, some identification of these follow:

1. *Clerical Biographical Inventory.* An interest and experience questionnaire keyed for clerical specialties, such as word knowledge and numerical operations. The same inventory is differentially keyed for other specialties as indicated below.
2. *Mechanical Biographical Inventory.* Keyed for specialties such as aircraft and engine maintenance.
3. *Crafts Biographical Inventory.*
4. *Equipment Operator Biographical Inventory.* Keyed for specialties such as construction machinery operation.
5. *Radio Operator Biographical Inventory.*
6. *Services Biographical Inventory.* Keyed for such specialties as food services.
7. *Electronics Technician Biographical Inventory.*
8. *Word Knowledge.* A multiple-choice vocabulary test.
9. *Arithmetic Reasoning.* Arithmetic problems.
10. *Dial and Table Reading.* Reading instrument dials and finding information by reading mathematical tables.
11. *Numerical Operations.*
12. *Aviation Information.*
13. *Background for Current Affairs.* Informational test.
14. *Electrical Information.*
15. *Mechanical Principles.* A mechanical comprehension test using pictures and diagrams.
16. *General Mechanics.* A test of practical mechanical information.
17. *Tool Functions.* An information test about the use of various tools.
18. *Speed of Identification.* A test of speed of identifying the silhouette of an object which has been rotated and imbedded in a group of similar silhouettes.
19. *Memory for Land Marks.* Test of memory for verbal designations of visual forms designed to test rote memory.
20. *Pattern Comprehension.* A test of understanding of relationships between solids and their flat projections.

*Subjects.*—The *S*s were 156 basic trainee airmen selected from the population at Lackland AFB and tested on the Complex Coordination Test. Printed tests scores were available for 149 of these *S*s.

*Procedure.*—The *S*s were trained for six 1-hr. periods. A rest of 15 min. was permitted between all periods except the third and fourth, where approximately 1 hr. was allowed for lunch. The number of matchings was recorded at 1-min. intervals for the first 10 min. of practice and at 2-min. intervals for the next 50 min. of practice. For the rest of the training session scores were recorded at 5-min. intervals.

## Results

Means for all scoring periods and means and standard deviations for each hour are presented in Table 1. The acquisition data stated in terms of matchings per min. are plotted in

TABLE 1

MEAN SETTINGS AND *SD*'S FOR 20-MIN. PERIODS FOR EACH HOUR

| Hour | 20-Min. Periods Within Hours | | | | | |
|---|---|---|---|---|---|---|
| | 1 | | 2 | | 3 | |
| | Mean | *SD* | Mean | *SD* | Mean | *SD* |
| 1 | 93.54 | 24.39 | 119.85 | 27.11 | 133.56 | 30.47 |
| 2 | 150.98 | 30.78 | 154.72 | 34.32 | 158.63 | 35.83 |
| 3 | 165.90 | 35.72 | 162.15 | 39.90 | 162.08 | 40.57 |
| Rest 1 hr. | | | | | | |
| 4 | 148.18 | 36.91 | 157.50 | 40.05 | 160.31 | 43.90 |
| 5 | 173.72 | 41.91 | 169.79 | 45.48 | 171.70 | 47.08 |
| 6 | 183.76 | 42.51 | 179.58 | 46.03 | 180.26 | 48.04 |

Fig. 1. Although the present study is not directly concerned with the characteristics of the acquisition curve for the complex coordination test, certain features of the curve are worthy of some comment. The curve, in general, rises throughout the training session. The early segments of the curve show rapid negative acceleration tending towards linearity.

Reminiscence apparently occurs following all 15-min. rest periods. Decrement is observed following the longer rest interval occurring during the lunch period. For the later periods, when reminiscence occurs following rest, later performance tends to drop to a level which remains relatively stable to the end of the hour period. There is a suggestion of final end spurt in performance for most hour periods.

In the case of the fourth hour, where initial decrement is observed, the performance curve *rises* to a relatively

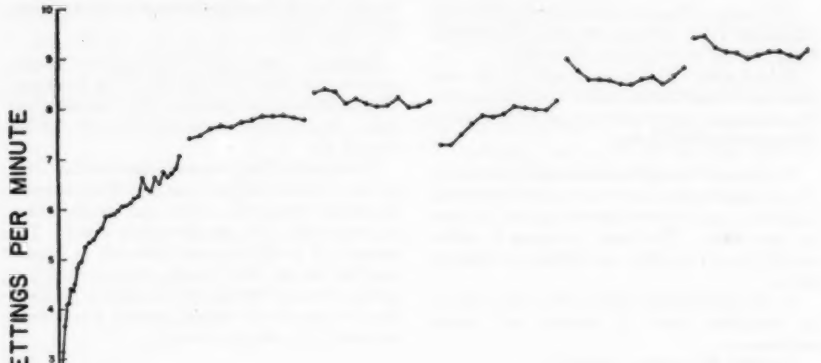


FIG. 1. Mean settings per minute plotted against cumulative practice time. The separations indicate the occurrence of rest periods.

stable level towards the end of the hour period. It is interesting to note that if the fourth-hour performance curve is rotated about its apparent asymptote, the curve will be seen to be very similar in shape to those for the fifth and sixth hours. This suggests that the conditions of testing employed established during each hour period a level of performance which was to some extent independent of initial performance in the hour. When reminiscence occurred, the readjustment was downward, whereas when initial decrement was observed, the adjustment was similar to that following reminiscence but in the opposite direction.

Table 2 contains correlations of the printed tests with selected part scores from the Complex Coordination Test. Most correlations are reduced significantly from the first 20 min. to the twelfth 20 min. of testing (Table 2). The mechanical experience and knowledge tests are affected more than the perceptual tests.

By the end of training only Speed of Identification, Dial and Table Reading, Numerical Operations, and the Clerical Biographical Inventory Key are significant at the 1% level of confidence. Arithmetic Reasoning, Memory for Landmarks and Radio Operator Biographical Inventory Key are significant at the 5% level.

Table 3 has correlations between the first 20-min. and last 20-min. scores for all hours. These 20-min. periods have been numbered in order of ap-

TABLE 2

CORRELATIONS WITH INDICATED 20-MIN. PERIOD SCORES ON THE COMPLEX COORDINATION TEST OF 20 VARIABLES FROM THE AIRMAN CLASSIFICATION BATTERY

(Decimal points omitted)

| Printed Test Variables | | | Correlations with 20-Min. Scores on the Complex Coordination Test | | | | Difference $r_{12} - r_1$ |
|---|---|---|---|---|---|---|---|
| Title | Mean | *SD* | 1st | 2nd | 7th | 12th | |
| *Biographical Inventory Keys* | | | | | | | |
| Clerical | 5.114 | 1.94 | 12 | 10 | 19 | 21*** | 09 |
| Mechanical | 4.926 | 1.87 | 40*** | 30 | 16 | 11 | −29*** |
| Crafts | 4.826 | 1.66 | 29*** | 25 | 12 | 07 | −22*** |
| Equipment Operator | 5.920 | 1.90 | 01 | 00 | −11 | −08 | −09 |
| Radio Operator | 5.007 | 1.75 | 28*** | 18 | 24 | 18* | −10 |
| Services | 5.383 | 1.98 | −18* | −18 | −15 | −14 | − 4 |
| Electronics | 4.879 | 1.81 | 30*** | 24 | 17 | 12 | −18** |
| *Test Variables* | | | | | | | |
| Word Knowledge | 6.430 | 1.75 | 19* | 13 | 11 | 10 | − 9 |
| Arithmetic Reasoning | 6.362 | 1.68 | 24*** | 20 | 17 | 16* | − 8 |
| Dial and Table Reading | 5.758 | 1.70 | 44*** | 36 | 27 | 27*** | −17** |
| Numerical Operations | 5.812 | 1.64 | 23*** | 26 | 24 | 25*** | 02 |
| Aviation Information | 5.926 | 1.75 | 31*** | 21 | 13 | 13 | −18** |
| Background for Current Affairs | 5.966 | 1.81 | 17* | 10 | 10 | 15 | −02 |
| Electrical Information | 6.262 | 1.87 | 42*** | 26 | 12 | 07 | −35*** |
| Mechanical Principles | 6.094 | 1.90 | 46*** | 30 | 21 | 14 | −32*** |
| General Mechanics | 6.309 | 2.12 | 39*** | 21 | 13 | 06 | −33*** |
| Tool Functions | 6.349 | 1.97 | 39*** | 24 | 09 | 02 | −37*** |
| Speed of Identification | 5.584 | 1.81 | 42*** | 36 | 29 | 27*** | −15* |
| Memory for Landmarks | 5.463 | 1.83 | 21*** | 21 | 12 | 20** | −01 |
| Pattern Comprehension | 6.617 | 1.79 | 32*** | 26 | 14 | 10 | −22*** |

* Significant at the 5% level.
** Significant at the 2% level.
*** Significant at the 1% level.

pearance, e.g., 12 is the last 20 min. in the sixth hour, 3 is the first 20 min. in the second hour. Two trends in the correlational data are readily observable. First, the $r$'s decrease as the distance in training time between the scoring periods increases. This is most easily seen in the correlations for the first period scores where $r_{1,2}$ is. 85 and $r_{1,12}$ is .58. The trend of decreasing correlation was so regular and consistent that it didn't seem necessary to apply a test of significance. There is one divergence from the general trend in the case of the $r$'s for the last scoring period before lunch with all following periods. Here there is a tendency for the $r$'s to increase and then decrease. The significance of this divergence will be discussed later.

The second trend in the data is an increase in the correlation between adjacent scoring periods with increases in the amount of practice preceding them. The $r$'s involved are those in the longest diagonal of the correlation table. Again there is a divergence from the main trend in the case of the $r$'s for the period immediately before lunch, accompanied in this case by a similar divergence in the case of $r$'s for the period immediately following lunch. The stability of the trends is shown by the fact that the difference between the $r$'s for the first and second 20-min. scores and the $r$'s between the fifth and sixth 20-min. scores is significant ($t = 3.36$, $df = 150$) although the difference between the fourth and fifth periods and the fifth and sixth periods is zero. The same result holds for the difference between the correlations for the sixth and seventh periods and the eleventh and twelfth periods ($t = 6.17$) while the difference between the $r$'s for the tenth and eleventh period and the eleventh and twelfth is not significant ($t = .964$).

The two main trends just noted are illustrated by the graph in Fig. 2 and the graph for the first scoring period $r$'s in Fig. 3. In Fig. 2, $r^2$ for adjacent scoring periods (1 vs. 2; 2 vs. 3; 3 vs. 4; . . . 11 vs. 12) are plotted against number of previous scoring periods for the first member of the pair (i.e., 1 vs. 2 against 0; 2 vs. 3 against 1; . . . 11 vs. 12 against 10). There is

TABLE 3

INTERCORRELATIONS BETWEEN SCORES BASED ON THE FIRST AND LAST 20 MIN. IN EACH HOUR OF TRAINING; IDENTIFIED BY ORDER OF OCCURRENCE ($N = 153$)

(Decimal points omitted)

| Scoring Periods | Hour | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 1st | | 2nd | | 3rd | | 4th | | 5th | | 6th | |
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
| 1 | | 85 | 76 | 68 | 67 | 62 | 61 | 58 | 61 | 56 | 62 | 58 |
| 2 | | | 91 | 86 | 82 | 80 | 73 | 75 | 78 | 74 | 77 | 75 |
| 3 | | | | 93 | 89 | 86 | 76 | 75 | 80 | 76 | 79 | 77 |
| 4 | | | | | 92 | 90 | 78 | 80 | 84 | 79 | 83 | 80 |
| 5 | | | | | | 92 | 80 | 79 | 82 | 78 | 84 | 78 |
| 6 | | | | | | | 81 | 83 | 87 | 85 | 86 | 83 |
| 7 | | | | | | | | 87 | 87 | 81 | 86 | 81 |
| 8 | | | | | | | | | 92 | 89 | 90 | 90 |
| 9 | | | | | | | | | | 94 | 93 | 90 |
| 10 | | | | | | | | | | | 93 | 90 |
| 11 | | | | | | | | | | | | 94 |

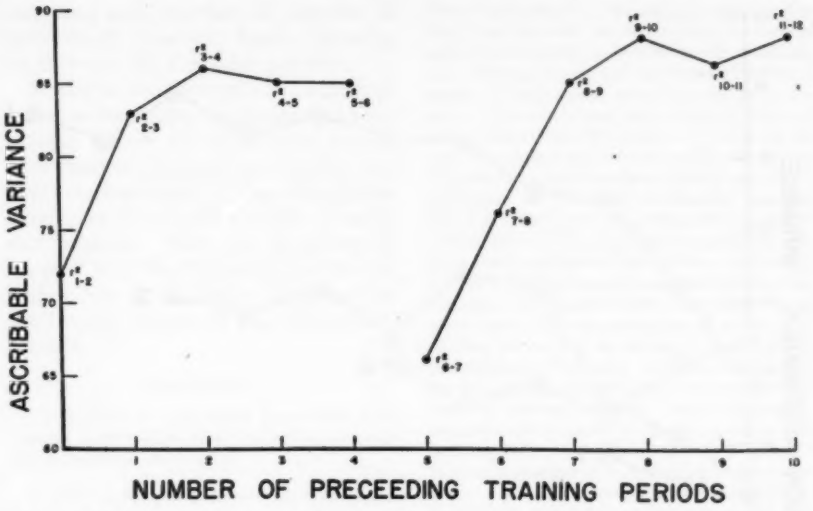


FIG. 2. Ascribable variance ($r^2$) between adjacent 20-min. periods plotted against number of periods preceding the first member of the pair, i.e., $r^2_{1,2}$; $r^2_{2,3}$; . . . ; $r^2_{11,12}$ against 0, 1, . . . , 10. The separation between $r^2_{5,6}$ and $r^2_{6,7}$ indicates 1-hr. rest.

a break in the curve at the period immediately before lunch. The graph is most readily interpreted in terms of per cent of variance in the scoring period which is ascribable to between-*S* differences observed in the immediately preceding scoring period. The per cent of ascribable variance increases from 71% for the first pair (1 vs. 2) to 85% for the fifth pair (5 vs. 6). The value for the sixth pair is 66%, rising to 89% for the last pair.

In Fig. 3 the curve with the longest extension is based on the correlations of the first-period scores with all following-period scores. The *r*'s have been squared, subtracted from 1.00 and the remainders plotted as per cents. These values represent the per cent of variance in the scoring period which cannot be ascribed to between-*S* variability observed during the first 20 min. of testing. The nonascribable variance in the second period is 29%, increasing to 69% in the tenth period and to 67% in the final period.

The reader will already have noted that both the curves just described have the general shape of acquisition curves. Further, the curve based on adjacent periods follows the general course of the performance curve in Fig. 1.

The two trends just described are most apparent and distinct. Other trends can be discerned although these are less obvious. For example, the curve in Fig. 3 (for the first scoring period with all subsequent periods) is based on decreases in ascribable variance as a function of training. The *S*s have had no previous experience with the test, since the *r*'s treated are for the first 20 min. of testing with later periods. The effect of increased training can be observed in the case where the period considered has been preceded by testing. This can be observed by considering either the cor-

FIG. 3. Nonascribable variance ($100-r^2 \cdot 100$) for the indicated 20-min. period and following periods plotted against amount of practice as indicated by the number of periods separating the two periods considered, e.g., $r_{1,2}$; $r_{1,3}$; . . . ; $r_{1,12}$ against 0, 1, . . . , 10. The numerals in the geometric forms identify the points on the graph.

relation changes in the rows or in the diagonals beginning at the upper-right corner of the correlation table.

In Fig. 3 nonascribable variance has been plotted against *amount* of practice separating the scoring periods for the first period, as already noted, and for the third, sixth, and ninth periods. It will be noted that the initial slope of the first-period curve is steeper than that for the later-period curves. The difference between the slopes for the third- and ninth-period curves is negligible and these curves are less widely separated from each other than the first-period curve is from the third-period curve. It is obvious that distance between scoring periods has less effect upon correlations as amount of previous training increases. Further, it is apparent that the trend towards decreasing correlation as a function of the intervening amount of practice is reduced after a relatively small amount of practice.

Considering the third-period curve alone, it will be observed that the interpolation of the rest period is accompanied by a sharp rise in the amount of nonascribable variance.

The curve for Period 6 involves correlations of periods following the lunch period with the scoring period immediately before lunch. Here the nonascribable variance is first higher than that for the other groups and then

decreases as a function of amount of intervening practice, finally showing an increase for the later periods.

It will be remembered that a 15-min. rest was interpolated between all hour periods except the third and fourth. The question arises as to whether any effect is observable in the correlations which can be ascribed to the 15-min. rest interval. The two main trends coupled with the manifest effect of the longer rest period make it difficult to discern any effects of the shorter rest periods.

## Discussion

Preliminary to considering hypotheses as to changes in correlations observed, the main trends may be summarized briefly.

1. The correlation between two test periods on a psychomotor test decreases as a function of the amount of practice separating the test periods. This trend can be observed in data reported by Greene (3), Houston (6), Melton (8), Reynolds (10), and Viteles (11). Heese (5) has reported initial and final score correlations for certain manual tests which were considerably lower than the reliabilities of the tests used.

2. The correlation between testing periods increases as a function of the amount of practice preceding the first test period in the particular pair considered. The sources cited above furnish supporting evidence for this with the present study affording perhaps the best illustration. Greene in his factor-analysis study found that with continued testing there were shifts in the size of factor loadings. The shift was usually toward a higher loading in one factor and lower loadings in the other factors.

3. The correlation of psychomotor tests with other tests, pencil and paper or psychomotor, has been shown to decrease as a function of training. This is one of the main results of the present study. It has been observed earlier by the writer (10) and by Viteles (11), and can be noted in data to be found in Melton (8) and Houston (6).

4. The trends can be accentuated by massing of training trials (10) and disrupted by the interpolation of a rest period of sufficient duration to produce a loss of proficiency in performance.

One simple hypothesis which can be advanced to explain the results is that the tests measure different things at different stages in training. Aside from simplicity this hypothesis has little else to recommend it. Actually, its seeming simplicity can be seen to derive from its lack of definition as to what is implied by the statement that different things are measured at different stages. If this means there is a change in the pattern of correlations with training, there is some evidence that this is true. It is true in the sense that certain variables that originally showed appreciable correlation show either no significant correlation or decreased correlation following training. This, however, is what the hypothesis should explain. The hypothesis may mean something above and beyond this. It may be assumed that each *S* brings to a task a set of unchangeable abilities. At different stages of learning the task draws upon different groupings of these. The abilities are entirely or relatively unchanged by performance of the task. It is difficult to evaluate this second formulation of the hypothesis in terms of the available data. Nor is the hypothesis particularly useful in explaining the specific features of the results which have been obtained. It is not consistent with the finding of the present study where the correlations with pencil and paper tests either decrease or remain the same with continued training. The only positive bit of evidence is the mergence of the low but significant *r* of the Clerical Biographical Inventory Key with later-period scores, although the change in correlation from the first to the twelfth scoring period does not reach significance. This may be only because the proper tests have not been chosen for study. Be that as it may, the results observed with the tests employed are most consistent with the notion that with continued testing the pool of abilities sampled is subjected to shrinkage, which is to say that with continued practice the test measures not different things but fewer things.

A second hypothesis, an inhibition hypothesis, has been suggested in an earlier report. It is commonly assumed that during training inhibition develops. This can be conceived of as reactive inhibition (7) internal inhibition (9) or inhibition arising from habit interference (4). Individual differences in the rate of development of such inhibition can be held to exist. As training continues these individual differences may make a greater contribution to the between-*S*s differences. The correlation of early trials with later trials is decreased as a consequence. As a second consequence the correlation between adjacent trials increases as inhibitory loadings approach higher, more stable levels. If individual differences in development of inhibition are assumed to be relatively unrelated to individual differences along other test dimensions, then with increasing training the correlation of a psychomotor test with other tests will decrease. There is nothing in the data from the present study

which is inconsistent with this hypothesis. However, data reported by Houston, based upon 30 *Ss*, can be interpreted as being inconsistent with it. Houston trained *Ss* on the Complex Coordination Test for 4 min. per day for 15 days. The same trends in the intercorrelations found in the present study can be observed in Houston's data. If growth of inhibition is the *sole* contributor to attentuation of the between-trials correlations, then attenuation should be considerably reduced by the spacing of 4-min. trials through 15 days. There were no massed-trial groups in the Houston study so it is impossible to evaluate the change in correlation with respect to distribution, but the distribution is wide enough that marked changes in intertrial correlations are not to be expected if growth of inhibition alone is the determinant of attenuation. It may be true, of course, that a 4-min. training period is not long enough to offset initial inhibitory effects produced by the 24-hr. interval. As a consequence individual differences in resistance to inhibitory effects might have been a very important source of variability.

A third hypothesis might be termed a learning hypothesis. According to this hypothesis *S* comes to a psychomotor task with a set of habits representing perhaps an interaction of experiential and certain structural and constitutional variables. The *S* uses these habits in performing the task. The *S*'s early trial performance is to a great extent a function of previous experience. To some extent, then, his performance reflects a number of abilities. As learning continues, to an increasing extent the performance of *S* depends upon his experience with the particular task employed. To an increasing extent, also, he learns the specific responses required by the task. The abilities or aptitudes associated with the restricted and specific habits being acquired are similarly progressively restricted in extent and generality as the habits employed become more specific.

With this hypothesis the correlational changes in the present study are readily explained. With increasing specificity in the habits acquired there is increased specificity of what is measured by the test. As a consequence there is decreased correlation between remote segments and increased correlation between adjacent segments. It is difficult, however, to explain the results obtained with massing and distributing of trials with a pure learning hypothesis.

The most adequate hypothesis, perhaps, is a general learning hypothesis. According to this hypothesis, which is essentially a combination of the learning and inhibition hypotheses just considered, there is increasing specificity of the habits acquired as a function of training. This tendency towards specificity can be assumed to arise because only a certain limited number of responses are most effective in performing the task. The responses being acquired become inhibited as a function of practice. Presumably there are between-*Ss* differences in susceptibility to inhibition, and these will make an increasing contribution to the variance observed. It may also be assumed that with increased learning many of the habits *S* brings to the task come to interfere with the responses being acquired. These more generalized response tendencies undergo extinction. Accordingly the unique character of what is being measured by the test is enhanced through elimination of these widely generalized responses. In the present study the decreased correlation between adjacent trials observed as an apparent consequence of the intervening longer rest period can be explained according to a general learning hypothesis as resulting from recovery of extinguished general responses. Supporting the learning hypothesis is the relationship between the general form of the acquisition curve and the curve for the increase of ascribable variance based on adjacent trials.

Learning theorists have given little, if any, attention to the problem of the relationship of individual differences to the principles of learning. This obviously makes it difficult to relate any specific theory to individual differences. It would seem to be consistent with most current learning theory that the laws of learning operate independently, to some degree at least, of individual differences; but that in a relatively complex learning situation the operation of the laws of learning is dependent, to some degree at least, upon fortuitous circumstances. These fortuitous circumstances may combine favorably for a person of mediocre ability and unfavorably for a person of greater ability. As a consequence the person of mediocre ability may learn more.

The results of the present study and of those previously mentioned suggest certain questions bearing upon many of the more important problems in the measurement of human behavior. There is insufficient information to permit answering these questions but trends are noticeable which indicate the kind of experimentation required. For this reason certain of these questions can be profitably noted here.

Assume a distribution of adequate measures of proficiency in the performance of a task learned, as the result of an intensive and extensive program of training supplemented by a long and continued period of application and refinement of the skill involved. What proportion of the variance of this distribution can be ascribed to measures based upon behavior sampling prior to the initiation of training? Is there any sizeable portion of the variance which is not predictable

in terms of any measures, other than those closely related to or nearly identical to the skill itself? Are presently available tests limited in their predictiveness to immediate criteria of proficiency and incapable of predicting terminal proficiency in a skill? Are efficient tests limited to those measuring widely generalizable responses such as are sampled by tests where understanding and manipulation of commonly used verbal and numerical symbols are involved? Is it possible to isolate variables which do not correlate to any n otable degree with immediate criteria but show increasing correlation with more remote criteria? Is it possible to isolate a limited number of variables which in the proper combination can account for all observed nonerror variance?

None of the above questions can be answered by consideration of the data of the present study. What can be stated on the basis of the present results is that increased training on a psychomotor test can be expected to decrease the proportion of the between-*Ss* variability which can be ascribed to variability on other test measures including measures from the same test taken at the beginning of training. On the other hand, with increased training psychomotor test performance becomes an increasingly efficient predictor of itself. For example, in the present study, between-*Ss* variability during the first 20 min. of training accounts for 33% of the variance in the scores for the last 20 min. of training. This contrasts with the predictability of the next to last 20-min. scoring period which accounts for 89% of the variance in the last 20 min. of training. Also pertinent to the questions raised above is the finding of the present study that particular characteristics of the acquisition curve can be directly related to the intercorrelational pattern. The data available are admittedly limited in number and restricted to a few tasks. But these data are consistent with the hypothesis that some portion of the between-*Ss* variance observed at advanced levels of proficiency in a task is not ascribable to variation along any other dimension other than that established by the task itself. From an experimental point of view, perhaps the most pressing question among those raised above is whether tests can be identified whose correlations with a psychomotor test will be greater at some later stage than they are at some earlier stage of training. Obviously, this is a necessary first step in the isolation of variables which in the proper combination could account for between-*Ss* variability at any stage of learning.

Turning, finally, to the printed test correlations with the final 20-min. scores, it is interesting to speculate as to the nature of what in final performance is predicted by the tests. The significant tests are Speed of Identification, Numerical Operations, Arithmetic Reasoning, Dial and Table Reading, Memory for Landmarks, and the Clerical and Radio Operator Biographical Inventory Keys. It seems safe to say that perceptual speed, numerical ability, and rote-learning ability are related to final performance in the present study. This would in turn suggest that some measure of speed of response is involved. Such a measure might be considered analogous to latency measures as these are employed in learning studies. In the case of numerical ability Coombs (1) has attempted to verify the hypothesis that what is involved in such ability is facility in the use of a set of highly stereotyped and highly overlearned set of habits and the results he obtained are consistent with this hypothesis. If the hypothesis is correct, then the consistent correlation of Numerical Operations with Complex Coordination may be considered to stem from common variance due to individual differences in facility in utilization of relatively well-learned habits. In this connection it is interesting to note that the Memory for Landmarks Test, which is designed to measure an aspect of learning ability, like Numerical Operations shows no attenuation with practice. This suggests the presence of some common learning ability which might prove measurable in terms of efficiency of utilization of well-learned responses.

## Summary

It is known that with increased practice of a psychomotor task two changes occur in the intercorrelations of trial scores. First, the correlation of early and remote trial scores decreases as the amount of practice separating them increases. Secondly, the correlation of adjacent trial scores increases with training.

The present study was designed to study these changes further and also to ascertain whether the correlation of a psychomotor test and a set of printed tests also decreases as a function of the amount of practice on the former.

Twenty printed tests from the Airman Classification Test Battery were given to the *Ss*. Subsequently they were given six 1-hr. practice periods on the Complex Coordinator. Fifteen minutes were given between all periods except the third and fourth, where an

hour was allowed for lunch. It was found that:

1. Correlation between two test periods decreased as a function of the amount of practice separating them. The distance between scoring periods was found to have less effect upon the correlations as the amount of training preceding the periods compared increased.

2. The correlation between test periods increased as a function of the amount of practice preceding the first period.

3. These trends are disrupted by the interpolation of a rest period of sufficient length to produce decrement in performance.

4. The correlations of the psychomotor test with most of the printed tests decreased as a function of training on the psychomotor test.

(Received January 22, 1952)

## References

1. Coombs, C. H. A factorial study of motor ability. *Psychometrika*, 1941, **6**, 161–189.
2. Dailey, J. T., Brokaw, L. D., & Lecznar, W. B. The development of the Airman Classification Test Battery. *Research Bulletin* 48–4. Barksdale AFB: Hq., Air Training Command, Nov., 1948.
3. Greene, E. B. An analysis of random and systematic changes with practice. *Psychometrika*, 1943, **8**, 37–52.
4. Guthrie, E. R. *The psychology of learning*. New York: Harper, 1935.
5. Heese, K. W. A general factor in improvement with practice. *Psychometrika*, 1942, **7**, 213–223.
6. Houston, R. C. An evaluation of the predictive properties of measures of variability of performance on three psychomotor tasks. Unpublished Ph.D. thesis, Univer. Maryland, 1950.
7. Hull, C. L. *Principles of behavior*. New York: D. Appleton-Century, 1943.
8. Melton, A. W. (Ed.) *Apparatus tests*. Report No. 4, Army Air Forces Aviation Psychology Program Research Reports. Washington: U. S. Govt. Print. Office, 1947.
9. Pavlov, I. P. *Conditioned reflexes* (trans. by J. V. Anrep). London: Oxford Univer. Press, 1927.
10. Reynolds, B. Correlations between two psychomotor tasks as a function of distribution of practice on the first. *J. exp. Psychol.*, 1952, **43**, 341–348.
11. Viteles, M. S. The influence of training on motor test performance. *J. exp. Psychol.*, 1933, **16**, 556–564.

# AN EXPERIMENTAL ANALYSIS OF THE ERRORS IN ROTE SERIAL LEARNING[1]

JAMES DEESE AND FREDERICK H. KRESSE

*The Johns Hopkins University*

The present experiment is an attempt to analyze the familiar bowed serial-position effect found in rote verbal learning in terms of the kinds of errors *Ss* make. It is, in this sense, an empirical approach to the problem of serial position in anticipation learning. The present experiment, however, differs from the previous experiments that make use of an analysis of intrusive errors in that, in the major experimental conditions, *Ss* were unpaced during learning. Together with the instructions given the *Ss*, this condition allowed every opportunity for *Ss* to verbalize any errors that they might have had a tendency to make during learning. Failures to respond were a result of an inability on the part of *Ss* to make a response with unlimited time. The *Ss* were not forced to guess, so that the number of times failure to respond occurred at each position could be tabulated.

## PROCEDURE

Three different experiments were performed, and the procedure for each of these will be described separately. The following general procedures apply to all of the experiments, however: All *Ss* were instructed to learn by the method of anticipation and to spell all of their responses. All *Ss* learned a preliminary list of six nonsense syllables consisting of two consonants with a vowel in between (hereafter called "vowel" syllables). These were presented to *Ss* by means of flash cards. This list was learned to a criterion of two consecutive errorless trials. The *E* used this opportunity to correct the learning technique of *S*. It was anticipated that the preliminary learning might give *E* an opportunity to discard unsatisfactory *Ss* at this point, but the occasion never arose. After a brief rest, *E* proceeded to the body of the experiment, which in every case consisted of learning, by the method of anticipation, a 12-item list of nonsense syllables. In Exp. I and II, a projection exposure device was used. Because of the limitation of alterations in the order of the film strips, however, in Exp. III a Gerbrands memory drum was used. All *Ss* were tested individually, and were instructed that *E* was interested in the kind of mistakes made during learning. In all experiments *Ss* were brought to a criterion of two consecutive correct trials.

*Experiment I.*—Ten *Ss* learned vowel syllables chosen from those with Glaze (1) association-values between 40% and 13%. The syllables were chosen so that no consonant appeared in either the first or third location more than once, and no vowel was duplicated more than twice. A 6-sec. interval was interpolated between each trial. The *Ss* were signaled the beginning of a trial by *E*. Each syllable appeared at the exposure window for a period of 2 sec., and there was a 2-sec. blank exposure between each syllable.

Two orders of list presentation were used. In the second order, the two halves of the list were exchanged and turned mirrorwise so that the syllables previously at the middle were now at the ends.

*Experiment II.*—Thirty *Ss* were used in this experiment. The procedure was exactly the same as in Exp. I except that *Ss* were not paced. The syllables were exposed in the window for 2 sec., as before, but the interval between syllables was prolonged until *S* responded or said that he could not think of the next syllable. The *Ss* were not forced to guess. The lists of Exp. I were used in this experiment.

*Experiment III.*—Twelve *Ss* learned lists of Witmer (4) consonant syllables of 13% to 42% association value which were presented on a memory drum. The order of the items in the list was different for every *S* (randomized in a latin-square design against *Ss*). The syllables were so devised that no consonant appeared in any one of the three locations more than once.

*Scoring errors.*—All responses were recorded and subsequently analyzed into three categories: (*a*) failures to respond, (*b*) errors or "intrusions" coming from other syllables within the list, and (*c*) overt errors not attributable to items within the list. The intralist intrusions were of two types, "complete intrusions" and "partial intru-

[1] The authors are indebted to Miss Elizabeth Ormond who gathered a portion of the data.

sions." Complete intrusions were complete spellings of syllables from other locations at an incorrect point. Partial intrusions were intrusions of two consecutive letters contained in syllables at other locations. Thus, for example, if *S* gave "Q–E–__" or "Q–E–T," and there was a syllable in the list, "Q–E–Z," this would be counted as a partial intrusion.

The validity of partial intrusions was checked by comparing their frequency at each position with complete intrusions, and they follow each other very closely. Because of necessary duplication in the vowel position in the vowel syllables, partial intrusions for these syllables were more equivocal than in the case of the all-consonant syllables. The results for the two cases were almost exactly the same, however (as can be inferred from an inspection of Fig. 2, 3, and 4). Therefore, the partial and complete intrusions were combined for the present analysis. In Exp. I the partial intrusions represented nearly 60% of the total intrusions, and in Exp. II and III about 40%. The higher figure for the first experiment is due to the fact that under conditions of pacing, *S*s frequently gave a blank for the last letter.

## Results

The results of Exp. I are presented in Fig. 1. A classical serial-position curve was found for total errors which was rather steep because of the comparatively rapid pacing. The intralist intrusions were of low frequency and bow shaped, as you can see in the bottom curve of Fig. 1. A striking fact appears, however: unlike the serial-position effect itself, the distribution of intralist intrusions is almost symmetrical. The frequency of such errors is just as great at the twelfth position as at the first. The highest frequency of such errors occurs at the seventh position and the next highest at the fourth. It cannot be unequivocally said that the distribution of intralist intrusions is perfectly symmetrical, but it is certainly more symmetrical than the distribution of total errors. It cannot be concluded that the remaining empirical sources of errors lead to the skewing effect, however, because no account is taken of implicit errors or errors that *S* almost made but couldn't because of the time. It is widely agreed, with data from many sources, that conflict between correct and incorrect responses produces long latencies of responses, so it was decided to repeat the experiment giving *S* unlimited opportunity to respond to each syllable.

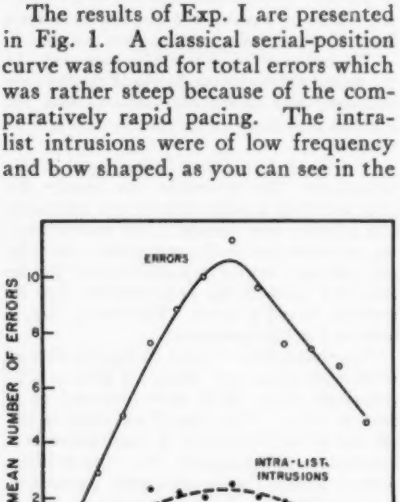


Fig. 1. Serial-position curve and frequency of intralist intrusions per syllable in Exp. I (paced learning of vowel syllables)

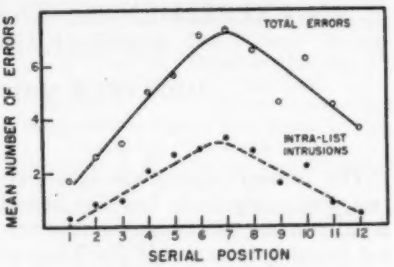


Fig. 2. Serial-position curve and frequency of intralist intrusions per syllable in Exp. II (unpaced learning of vowel syllables)

The serial-position curve of total errors for Exp. II is shown in Fig. 2. Again we have the familiar curve, but this time much reduced in slope because of the longer average time between syllables. The intralist intrusions are about of the same order of magnitude as in Exp. I, but they account for a

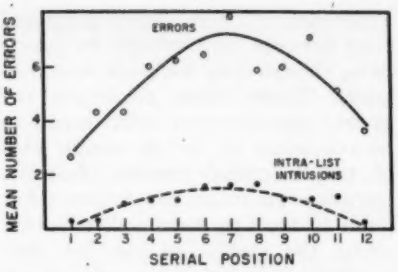


FIG. 3. Serial-position curve and frequency of intralist intrusions per syllable in Exp. III (unpaced learning of consonant syllables)

much larger proportion of errors (Fig. 2). Again the striking fact is that the the serial-position effect is, as usual, skewed, but the intralist intrusions are almost symmetrical. The slope of the curve for the last half of the list appears to be a little steeper than that for the first half, but it is not nearly so skewed as the curve for total errors. This time, however, if we now tabulate the remaining two categories of errors, we come up with an additional interesting fact. The failure-of-response errors are very low at the beginning of the list, rise rapidly through the first half of the list, and are approximately level through the second half of the list. This can be seen in Fig. 4. The overt errors coming from without the list (source unknown, perhaps "guessing," perhaps from the practice list) show no consistent pattern as a function of serial position. The mean frequencies of such errors per position, in order, are as follows: .76, 1.24, 1.34, 1.13, 1.43, 1.23, 1.33, .83, .46, .87, .90, .90.

Thus it appears that the skewed serial-position effect found in this experiment can be accounted for largely by two empirical factors: (*a*) intralist intrusions that are almost symmetrically bowed, and (*b*) failures to respond that increase through the first half of the list and reach a plateau during the second half. There are several features about this experiment that left us with the desire to pursue the matter further, however. The scoring of partial intrusions was not altogether unequivocal (because of necessary duplication of vowels in a 12-item list) and the list was presented in two orders, which did not clearly eliminate any confounding factors due to intrinsic difficulty of syllables. Therefore, the experiment was repeated using all-consonant syllables and a completely randomized order of items in lists.

The serial-position curve of Exp. III is again the classical curve (Fig. 3). Also, as in the case of the two previous experiments, the intralist intrusions are almost symmetrical about the middle. This distribution is more heavily skewed towards the end of the list than those of Exp. II, however. Also, as in Exp. II, the failure-of-response errors increase through the first half of the list and are approximately level through the last half (Fig. 4). The principal difference between this experiment and Exp. II is in the larger frequency of extralist overt errors in Exp. III, although they again show little consistent pattern. The mean frequencies per syllable are as follows: 2.08, 1.83, 3.00, 2.66, 3.50, 1.92, 2.16, 3.00, 1.07, 1.00. The larger frequency of these errors is responsible for making

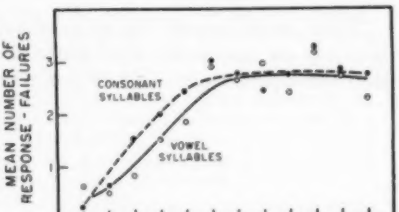


FIG. 4. Frequency of failure-of-response errors per position in Exp. II and III (unpaced learning of vowel and consonant syllables)

the serial-position curve itself flatter in Exp. III than in Exp. II. The bowness and skewness of the curve are entirely accounted for, however, by the intralist errors and the failures of response.

## Discussion

It appears that there are two empirical factors responsible for the skewed serial-position curve obtained with unlimited recall times. One is the almost symmetrical distribution of intralist intrusions and the other is the growth of failures to respond through the first half of the list to reach an asymptote during the second half. These facts suggest a two-factor theory of the serial-position curve in unpaced learning. A factor of associative inhibition would be indicated by the intralist intrusions, and a second factor, that might be identified with either the notion of response-induced inhibition or the immediate memory span, would be indicated by the failure-to-respond curves.

The serial-position curve obtained with paced learning appears to be more complicated. Failures of response, instead of growing in frequency through the first half of the list, show a bow-shaped, skewed curve. Further experimental work will be necessary to establish the reason for this difference.

## Summary

Three experiments on rote verbal learning by the method of anticipation were performed. In two of the experiments *S*s were allowed unlimited time between each syllable to anticipate, though they were not forced to guess. Under these conditions the skewed serial-position effect seems to be accounted for by the simple sum of two empirical factors. Intralist intrusions (mistakes based upon other items in the list) form a distribution about the middle of the list that is almost symmetrical. Failure-of-response errors are low at the beginning of the list, mount rapidly through the first half of the list, and reach a somewhat variable asymptote during the second half of the list. These two sources of errors, summed together, produce the classical serial-position curve. Response errors coming from the outside of the list form no consistent pattern as a function of serial position. By and large, they serve merely to flatten the serial-position curve.

(Received February 2, 1952)

## References

1. Glaze, J. A. The association value of nonsense syllables. *J. genet. Psychol.*, 1928, **35**, 255–267.
2. Hull, C. L. The conflicting psychologies of learning—a way out. *Psychol. Rev.*, 1935, **42**, 491–516.
3. Hull, C. L., Hovland, C. I., Ross, R. T., Hall, M., Perkins, D. T., & Fitch, F. B. *Mathematico-deductive theory of rote learning.* New Haven: Yale Univer. Press, 1940.
4. Witmer, L. R. The association value of three-place consonant syllables. *J. genet. Psychol.*, 1935, **47**, 337–359.

# EFFECT OF INTERTRIAL INTERVAL DURING ACQUISITION ON EXTINCTION OF THE CONDITIONED EYELID RESPONSE FOLLOWING PARTIAL REINFORCEMENT [1]

DAVID A. GRANT, LOWELL M. SCHIPPER, AND BRUCE M. ROSS

*University of Wisconsin*

The primary problem was to investigate the effect of the distribution of preceding reinforced trials on the extinction of the conditioned eyelid response following 100% and 50% random reinforcement. Over ten years ago, Humphreys (**6, 7, 8**) reported that 50% random reinforcement produced an unusual resistance to extinction in the conditioned eyelid response. This will be called the Humphreys random-reinforcement effect. Since then, V. Sheffield (**12, 13**) has shown that in a rat runway problem the heightened resistance to extinction following 50% random reinforcement does not occur when the training trials are spaced. In contrast, Grant, Hornseth, and Hake (**3**), investigating the "verbal conditioning" of human *S*s, observed that the distribution of the training trials made no difference in the Humphreys random-reinforcement effect. It is not particularly surprising that different response systems in different species should react differently to massing and spacing of training. It has been claimed, however, that principles which suffice to explain rat runway behavior (**12,** p. 523) are adequate for the explanation of Humphreys' original findings (**6, 7, 8**). It is therefore advisable to make a direct test of the implication of Sheffield's interpretation that following 50% *spaced* random reinforcement there will be some diminution of the Humphreys random-reinforcement effect in human eyelid conditioning.

It is fairly obvious that for an adequate comparison three factors: distribution of reinforcement, percentage of reinforced trials, and distribution of extinction must be varied in the experimental design (**3, 12, 13**). Analysis of such an experiment should yield further evidence: (*a*) on the effect of distribution of extinction trials on the course of extinction; (*b*) on the effect of distribution of reinforced trials on acquisition; (*c*) on the effect of distribution of acquisition trials upon the course of extinction; and (*d*) on the interactions of these factors. Several of these relationships have been investigated previously (**1, 10, 11, 14**).

## Apparatus and Procedure

*Apparatus.*—The *S* sat in a straight-backed chair facing a large rectangular enclosure placed on top of a table. The sides of the box were painted a flat black and the back consisted of a white matte surface. The enclosure was 95 cm. wide, 66 cm. high, and 122 cm. deep. About 80 cm. in front of *S*'s eyes and slightly below his line of level vision was a black rectangular box with a 10-cm. circular milk-glass window in it. The CS consisted of the illumination of this window. To the right of *S* was a 60-cm. square, one-way vision window which appeared to *S* as a mirror. Also present within the enclosure was a desk microphone and three 20-w. fluorescent lamps which were kept lighted at all times. The *S* wore a softball catcher's mask with most of the protective bars removed so that it retained only the pad which rested against his head and the outlined bars holding the stimulus air-jet and

[1] This research was supported in part by the Research Committee of the Graduate School of the University of Wisconsin with funds appropriated by the State Legislature.

the microtorque potentiometer[2] which served to pick up *S*'s eyelid movements. The microtorque potentiometer was coupled to the eyelid as follows: A hypodermic tube was cemented to the shaft of the potentiometer and a length of piano wire was passed through the tubing so that it slid smoothly. A small piece of thin aluminum was attached to the end of the piano wire. The aluminum strip was attached to *S*'s eyelid by means of Scotch Tape so that any movement of *S*'s eyelid would be transmitted to the shaft of the potentiometer. The most significant force acting against *S*'s eye was the frictional torque of the potentiometer which amounted to about .004 in.-oz. The timing and recording equipment was electronic. The signal from the potentiometer was amplified by a DC amplifier and recorded by means of an inkwriter. On the recording paper the onset and termination of CS and UCS were also indicated. The amplification was nonlinear so that frequency data only were analyzed in the present experiment.

The CS consisted of a light which lasted .75 sec. On reinforced trials the UCS was a corneal air puff administered .45 sec. after the onset of the light. The *S*s were run under light-adapted conditions to minimize the presence of sensitized $\beta$ responses (4). To retain light adaptation, the prestimulus field was maintained at a brightness of 4.4 millilamberts. The brightness of the CS was 121 millilamberts. The puff was adequate to elicit a sharp reflex blink. The air puff was applied to the left eye; recorded eyelid movements were from the right eye.

*Procedure.*—All *S*s were given 60 training trials on the first day of the experiment. On the second day of the experiment, 30 training trials were given, followed by 30 extinction trials. On massed trials, the intertrial interval was between 5 and 15 sec. with an average of about 10 sec. On spaced trials, the intertrial interval was between 40 and 50 sec. with an average of about 45 sec. For one-half of the *S*s, the training procedure involved 100% reinforcement; for the other half of the *S*s, the training procedure involved 50% random reinforcement. Nine different random reinforcement orders were used in a random order with different *S*s.

*Experimental design.*—The design of the experiment was identical to that of Sheffield (12, 13) and Grant, Hornseth, and Hake (3). All combinations of three factors, (*a*) massed or spaced training trials, (*b*) massed or spaced extinction trials, and (*c*) 100% or 50% reinforcement during training, were used, making eight procedures in all. Nine *S*s were used in each of the eight groups. Thus, 64 *df* from replication were available to estimate the error in the experiment.

*Subjects.*—The *S*s were men and women students from introductory classes in psychology at the University of Wisconsin. They were assigned at random to the eight experimental groups. The *S*s were instructed to maintain a passive attitude toward the experiment and not to influence in any way their natural reactions to the stimuli.

## Results

For purposes of the present report, all eyelid responses with latencies between 246 msec. and 445 msec. were called CR's. This latency range effectively eliminates the $\alpha$ reflex to light, the $\beta$ response, and reflexes to the puff. It is somewhat more restricted than latency ranges used in previous experiments in this laboratory.

In Fig. 1 are presented the results during training trials. Here, the per cent of CR's is plotted against successive six-trial blocks during the training period for all eight groups. Inasmuch as the massing and spacing of extinction trials makes no difference during acquisition, the eight groups have been combined into four as indicated in the figure. Thus, each point in the figure is based upon 18 *S*s, six trials each, or a total of 108 trials. In a general way, it can be seen that the 100% training schedule results in more CR's during the acquisition than does the 50% training schedule. This is particularly clear on the training trials of Day 2. On Day 1, spaced reinforcement appears to be more effective than massed reinforcement, but this relationship does not hold for the training trials on Day 2.

The analysis of variance summary table for acquisition and extinction results is given in Table 1. The acquisition analysis appears in the left-hand column of the table and the extinction analysis appears in the right-hand column. From the summary table, it

[2] The use of the microtorque potentiometer was suggested to the senior author by Prof. K. W. Spence of the State University of Iowa.

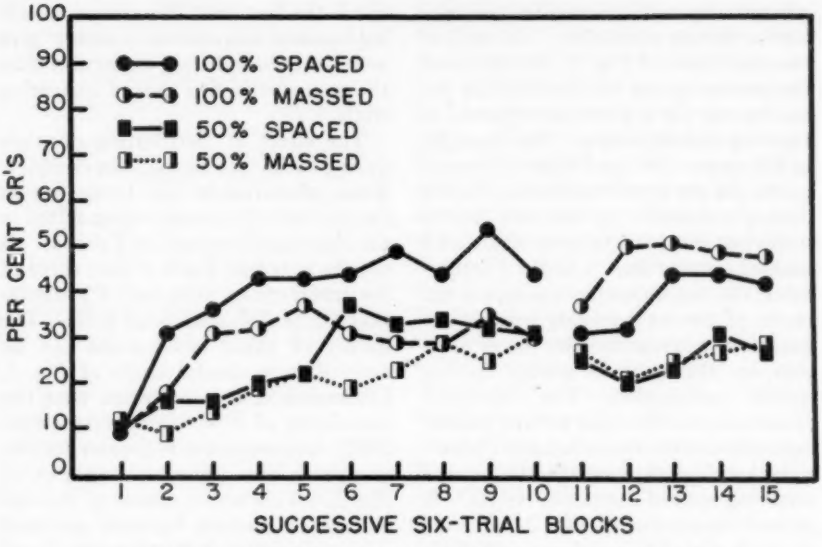


FIG. 1. Percentage of CR's plotted against successive six-trial blocks during acquisition. Percentage of reinforced trials and distribution of trials are the parameters.

appears that only the percentage of reinforced trials affects significantly the per cent frequency of CR's during training. The $F$ of 11.12 is significant at the 1% level. The $F$ of 1.05 indicates that the distribution of training trials did not have a significant effect upon the over-all frequency of CR's during acquisition.

The critical extinction data of the experiment are presented in Fig. 2. Here the per cent frequency of CR's is

TABLE 1

ANALYSIS OF VARIANCE SUMMARIES FOR THE TOTAL NUMBER OF CR'S DURING (A) ACQUISITION AND (B) EXTINCTION

| Source of Variation | (A) Acquisition Trials 1–90 | | (B) Extinction Trials 91–120 | |
|---|---|---|---|---|
| | *df* | *F* | *df* | *F* |
| Partial vs. 100% reinforcement | 1 | 11.12* | 1 | 29.76** |
| Distribution during acquisition | 1 | 1.05 | 1 | 7.44* |
| Distribution during extinction | | | 1 | 1/1.08 |
| Reinforcement × Acquisition | 1 | 1/19.49 | 1 | 1/2.43 |
| Reinforcement × Extinction | | | 1 | 21.65** |
| Acquisition × Extinction | | | 1 | 59.32** |
| Reinforcement × Acquisition × Extinction | | | 1 | 28.04** |
| Error (replication) | 68 | (227.69)*** | 64 | (2.16)*** |

* Significant at the 1% confidence level.
** Significant at the .1% confidence level.
*** Error estimate for the column.

plotted against successive six-trial blocks during extinction. In each of the four plots of Fig. 2, the effects of the percentage of reinforced trials are contrasted for a given distribution of training and extinction. For example, in the upper-left-hand figure are compared the per cent frequencies of CR's during extinction for the two groups receiving massed training trials and massed extinction trials. Furthermore, the two upper plots compare the result of groups receiving massed acquisition, whereas the two lower plots compare the result of groups having spaced acquisition. The left-hand plots compare the four groups receiving massed extinction trials and the two right-hand plots compare the groups receiving spaced extinction trials. A general inspection of Fig. 2 indicates that following 50% reinforcement there is, over-all, more resistance to extinction than in the groups following 100% reinforcement. Less readily apparent is the fact that the groups receiving massed acquisition training give more responses during extinction than the groups receiving spaced acquistion trials.

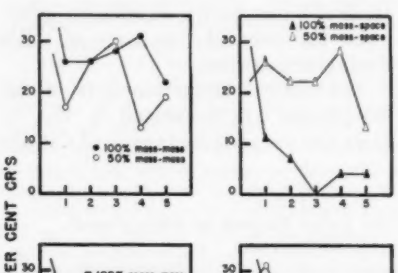


FIG. 2. Percentage of CR's plotted against successive six-trial blocks during extinction. In each graph curves are given following 100% and 50% random-reinforcement training schedules for two groups with identical distribution of trials during acquisition and extinction.

The effect of distribution of trials during extinction seems to be complex. These observations are borne out in the analysis of variance summarized in the right-hand column of Table 1. It will be noted in Table 1 that three of the interactions produced *F*'s significant at the .1% confidence level. The nature of these interactions can be derived from careful study of Fig. 2. For example, it is apparent that the superiority of 50% reinforcement over 100% reinforcement is greater for the groups in the right-hand graphs of Fig. 2. This is the source of the significant interaction between per cent reinforcement and the distribution of extinction trials. It can also be seen that extinction following massed reinforcement trials is most rapid if the extinction trials are spaced and the extinction following spaced reinforcement is most rapid with a massed schedule. In other words, taking the results of all groups together, changing the distribution of trials when the transition is made from training to extinction results in a more rapid extinction of the CR's. This fact is the basis of the significant interaction between distribution of trials during acquisition and distribution of trials during extinction. The nature of the three-factor interaction is somewhat harder to detect in Fig. 2. It will be noted, however, that the Humphreys effect, that is, the superiority of the 50% group over the 100% groups is more marked when there is a shift in the distribution of training and extinction trials. That is, if the graphs are compared diagonally, the upper left-hand and lower right-hand graphs show less Humphreys effect than do

TABLE 2

AMOUNT OF HUMPHREYS EFFECT OBTAINED WITH MASSING AND SPACING OF REINFORCEMENT AND EXTINCTION TRIALS*

| Reinforcement Schedule | Extinction Schedule | | Totals |
|---|---|---|---|
| | Massed | Spaced | |
| Massed | −16 | 46 | 30 |
| Spaced | 21 | 17 | 38 |
| Totals | 5 | 63 | 68 |

* Scores equal the number of CR's during extinction following 50% random reinforcement minus the number of CR's during extinction following 100% reinforcement.

the lower left-hand and upper-right-hand graphs.

The nature of the interaction effect is illustrated somewhat more clearly in Table 2. Here, a score has been obtained for each pair of groups in Fig. 2 by subtracting the number of CR's during extinction in the 100% group from the number of CR's during extinction in the 50% group for each of the four plots of Fig. 2. Thus, there were 16 more CR's given following 100% reinforcement than there were following 50% reinforcement where both the training and extinction trials were massed. There were 46 more CR's following 50% training than following 100% training when there was a transition from massing to spacing. There were 21 more CR's given during extinction in the 50% group than in the 100% group when training was spaced and extinction was massed, and there were 17 more CR's given in the 50% group than in the 100% group when both the training and the extinction trials were spaced. Table 2 is organized like any orthogonal analysis of variance table. The rows represent differences due to distribution of training trials, and the columns represent differences due to distribution of extinction trials. The difference between the diagonal sums represents the three-factor interaction. The marginal totals of 30 and 38 for the rows show that there is very little over-all difference in random-reinforcement effect caused by variation in distribution of training trials. The column totals show that the distribution of the extinction schedule is quite important in that a greater Humphreys effect is observable following spaced extinction trials. If the diagonals are summed algebraically, that is, subtracting 16 from 17 for a total of 1, compared with a sum of 67 for the other diagonal, the difference between +1 and +67 is the three-factor interaction effect. This shows that when the extinction schedule involves a change in the distribution of trials from that used in training, the Humphreys random-reinforcement effect becomes more markedly apparent.

## DISCUSSION

*Effect of distribution of training trials.*—One important finding of the present experiment was the failure of the phenomenon of heightened resistance to extinction following 50% random reinforcement to be affected by the intertrial interval during the training schedule. The experimental design, with 64 *df* in the error term, was powerful enough (2) to detect very slight differences. In regard to the degree of massing and spacing, the 10-sec. average interval between trials on spaced training and extinction was satisfactory, inasmuch as Calvin's data (1) showed little effect for spacing for intertrial intervals less than 10 sec. The average intertrial interval of 45 sec. seemed adequate enough in view of Calvin's data, but it might have been more satisfactory if distinct differences had been produced during acquisition. All told, however, the variation in spacing was great enough to produce significant interaction, so that the failure of distribution of training trials during acquisition to have an effect probably means that such an effect, if it exists, is not great for the conditioned-eyelid response.

In any case, it seems clear that the explanation of Sheffield is not sufficient, as claimed (12, p. 523), to account generally for the Humphreys random-reinforcement phenomenon. Just as an earlier experiment (3) showed that the Sheffield stimulus-pattern interpretation was not adequate for the Humphreys verbal-conditioning phenom-

ena, this study shows that the stimulus-pattern explanation can hardly account for the details of the eyelid-conditioning results. It must be pointed out that a similar interpretation proposed by Hake and Grant (5) is equally inadequate.

The difficulties of these interpretations can be adjusted to some extent by hypothesizing that the verbal mechanism of human *Ss* may result in relatively long aftereffects of stimulation. Even more drastic revision is required, however, in order that these theories may cope with the type of interactions found in this experiment as discussed in the next section.

*Effect of change of temporal distribution of training and extinction trials.*—Changing the distribution of trials at the time the extinction procedure begins increases the discrepancy between extinction curves of the 50% and 100% group. In other words, shifting from massed reinforcement to spaced extinction or from spaced reinforcement to massed extinction heightens the Humphreys random-reinforcement phenomenon. This finding is difficult to account for in terms of any presently held interpretation of the effect of partial reinforcement unless supplementary concepts are introduced. Not only is this three-factor interaction difficult to explain in terms of Sheffield's interpretation and in terms of the Hake and Grant interpretation, but it is also difficult to interpret it in terms of any of the possible explanations proposed by Mowrer and Jones (9) which depend fundamentally, like that of Humphreys, upon the discrimination of a change in schedule when extinction supersedes training procedures. In general, a shift in the distribution of trials should heighten the discriminability of the change from the reinforcement procedure. Certainly, the shift in distribution of trials markedly facilitates extinction for the 100% groups. If the greater resistance to extinction following partial reinforcement depends upon the greater similarity of extinction procedure to partial-reinforcement procedures, then anything which would heighten the discriminability of the change in schedule might well reduce the partial-reinforcement phenomenon. This is clearly not so, as indicated by the present result. In fact, the results show quite the opposite phenomenon.

The interaction of the acquisition and extinction distribution schedules with the percentage of reinforcement is caused chiefly by the marked drop in the 100% group where massed extinction follows spaced training or where spaced extinction follows massed training. This observation makes possible an a posteriori interpretation of the unforeseen findings. Perhaps with 100% reinforcement the simplicity of the training schedule permits *S* to attend to the distribution of trials more closely than is possible in the 50% reinforcement schedule. In other words, the complexity of the 50% schedule may distract *S* from the temporal interval between trials. It seems reasonable that the 50% schedule would therefore reduce the discriminability of the changes in intertrial interval between acquisition and extinction, so that the changes would not stand out as a cue to the fact that extinction had begun. This is admittedly an ex post facto interpretation which requires further test. Some supplementary interpretive concepts such as this are clearly required to account for the finding.

*Some relationships between human conditioning and rat behavior in the runway.*—The fact that the learning of rats in a runway (12, 13), human eyelid conditioning, and human "verbal conditioning" (3) have been investigated using identical experimental design gives an opportunity for a brief review of the differences to be observed with these species in three different types of learning situation. In neither human "verbal conditioning" nor in human eyelid conditioning does the distribution of training trials affect the Humphreys random-reinforcement phenomenon, whereas this interaction is found in rat runway behavior. Changing distribution of training and extinction trials affects the Humphreys phenomenon only in the conditioned eyelid response. Aside from these functional differences which have been emphasized in earlier discussion, the following may be noted: (*a*) partial reinforcement reduced the frequency of responses during acquisition in eyelid conditioning and "verbal conditioning" but did not affect runway behavior; (*b*) distribution of reinforced trials seems to affect eyelid conditioning only;[3] (*c*) only in rat runway behavior did distribution of trials clearly affect extinction; and (*d*) in eyelid conditioning spaced extinction trials emphasized the Humphreys effect, while massed extinction trials served the same function in runway behavior.

The foregoing summary of the differences in these three learning situations points up the fact that in theorizing about learning, care must be taken to restrict the generalizations obtained from one experiment to the kind of *S* and the kind of learning situation used in that experiment only. It is probable that if identical experimental designs can be used in a number of learning situations with human and other *Ss*, the common elements in the learning which do not depend upon the particular learning task or the particular species may stand out and be the basis for the generalized learning theory of the future,

[3] The present study provides no statistically significant proof of this, but Calvin (1), three of Reynolds' four groups (10, 11), and Spence and Norris (14) show that massing of training reduces the frequency of CR's during acquisition.

and behavior that is a function of the species and task may be attributed to those factors rather than to learning per se.

## Summary and Conclusion

The effect of temporal interval between training trials and between extinction trials on the heightened resistance to extinction brought about by partial reinforcement was investigated in an experiment with 72 human *Ss*. The conditioned eyelid response to light based upon the reflex to corneal air puff was used under standard light-adapted conditions. Ninety training trials on Days 1 and 2 were used, followed by 30 extinction trials on Day 2. Massing and spacing of training trials, 50% random reinforcement, and 100% reinforcement, and massing and spacing of extinction trials were the three variables in a trifactorial design. The frequency of CR's during acquisition and extinction was analyzed by the analysis of variance, and the principal findings were as follows:

1. During acquisition, there were significantly more CR's in the 100% reinforced group than the 50% reinforced group.

2. Significantly more resistance to extinction was found following partial reinforcement and following massed reinforcement than was the case following 100% reinforcement and spaced reinforcement, respectively.

3. The heightened resistance to extinction following partial reinforcement was significantly greater for spaced than for massed extinction trials.

4. Following spaced acquisition, massed extinction was significantly more rapid, and following massed acquisition, spaced extinction was significantly more rapid.

5. The heightening of the resistance to extinction following partial reinforcement was significantly greater in massed extinction if the reinforcement trials had been spaced and in spaced extinction if the reinforcement trials had been massed.

6. It was pointed out that interpretations of these results in terms of stimulus patterns following Sheffield, or in terms of expectancy following Humphreys, or in terms of discrimination of the extinction schedule from the training schedule following Mowrer and Jones or Hake and Grant were not adequate without at least two further concepts; i.e., one which might concern the role of verbalization in human *Ss*, and the other which would attempt to account for the second-order interaction of variables mentioned in the preceding conclusion.

7. The differences between human eyelid conditioning, human "verbal conditioning," and rat runway behavior were summarized in terms of the variables of distribution of training and extinction trials and partial reinforcement. It was pointed out that these differences preclude broad generalizations, but they may point to common characteristics of learning in a few situations.

(Received February 8, 1952)

## References

1. Calvin, J. S. Decremental factors in conditioned-response learning. Unpublished Ph.D. thesis, Yale Univer., 1939, cited in E. R. Hilgard and D. G. Marquis, *Conditioning and learning*. New York: Appleton-Century, 1940, p. 149.
2. Ferris, C. D., Grubbs, F. E., & Weaver, C. L. Operating characteristics for the common statistical tests of significance. *Ann. math. Statist.*, 1946, **17**, 178–197.
3. Grant, D. A., Hornseth, J. P., & Hake, H. W. The influence of the intertrial interval on the Humphreys' "random reinforcement" effect during the extinction of a verbal response. *J. exp. Psychol.*, 1950, **40**, 609–612.

4. Grant, D. A., & Norris, E. B. Eyelid conditioning as influenced by the presence of sensitized beta-responses. *J. exp. Psychol.*, 1947, **37**, 423–438.
5. Hake, H. W., & Grant, D. A. Resistance to extinction and the pattern of reinforcement: II. Effects of successive alternation of blocks of reinforced trials upon the conditioned eyelid response to light. *J. exp. Psychol.*, 1951, **41**, 216–220.
6. Humphreys, L. G. Acquisition and extinction of verbal expectations in a situation analogous to conditioning. *J. exp. Psychol.*, 1939, **25**, 294–301.
7. Humphreys, L. G. The effect of random alteration of reinforcement on the acquisition and extinction of conditioned eyelid reactions. *J. exp. Psychol.*, 1939, **25**, 141–158.
8. Humphreys, L. G. Extinction of conditioned psycho-galvanic responses following two conditions of reinforcement. *J. exp. Psychol.*, 1940, **27**, 71–75.
9. Mowrer, O. H., & Jones, H. Habit strength as a function of the pattern of reinforcement. *J. exp. Psychol.*, 1945, **35**, 293–311.
10. Reynolds, B. The acquisition of a trace conditioned response as a function of the magnitude of the stimulus trace. *J. exp. Psychol.*, 1945, **35**, 15–30.
11. Reynolds, B. Extinction of trace conditioned responses as a function of the spacing of trials during the acquisition and extinction series. *J. exp. Psychol.*, 1945, **35**, 81–95.
12. Sheffield, V. F. Extinction as a function of partial reinforcement and distribution of practice. *J. exp. Psychol.*, 1949, **39**, 511–526.
13. Sheffield, V. F. Resistance to extinction as a function of the distribution of extinction trials. *J. exp. Psychol.*, 1950, **40**, 305–313.
14. Spence, K. W., & Norris, E. B. Eyelid conditioning as a function of the intertrial interval. *J. exp. Psychol.*, 1950, **40**, 716–720.

# COMPARISONS OF FORWARD, SIMULTANEOUS, BACKWARD, AND PSEUDO-CONDITIONING

M. E. FITZWATER AND MARVIN N. REISMAN

*Bowling Green State University*

Several investigations have been conducted to determine the effect of various temporal relationships between conditioned and unconditioned stimuli upon the establishment of a conditioned response. Wolfle (**9**), Bernstein (**1**) and others have observed that conditioning occurs most readily when the CS precedes the US by approximately .5 sec. They also have reported CR's when employing simultaneous and backward modes of stimulus presentation, but in smaller magnitude than forward. Experimental findings have led Wenrick to conclude, "The greatest amount of congruence seems to be with reference to the inadequacy of backward conditioning, even though some report it possible, it is unstable and evanescent" (**7**, p. 533). Porter (**4**) has also questioned the adequacy of backward conditioning.

Wickens and Wickens (**8**) and Harlow (**2**) have observed apparent conditioned responses to the CS on test trials, even though the CS and US were never contiguously paired—the training trial being the presentation of the US alone. The occurrence of CR's to unpaired CS's has been termed *pseudo-conditioning*. Indications of simultaneous and backward CR's were found by Spooner and Kellogg (**5**). However, on the basis of extreme differences in the shape of the forward-conditioning curve as contrasted with the backward- and simultaneous-conditioning curves, they concluded that "two distinct processes are represented—one by forward conditioning and the other by simultaneous and backward conditioning" (**5**, p. 329). Moreover, they noted the possibility that the backward and simultaneous modes of presentation might be special kinds of pseudo-conditioned responses. This could not be stated with any degree of certainty as they had no index of pseudo-conditioning for comparison.

The present investigation was designed to determine the relative incidence of conditioning under forward, backward, and simultaneous conditions when compared with pseudo-conditioning results.

## Apparatus and Procedure

*Apparatus.*—In all groups an instrumental finger withdrawal to shock situation was used. A 1024-cycle audiometer tone was presented to *S* through headphones as the CS. An AC shock of variable intensity was the US and was adjusted to the minimal intensity necessary for the consistent withdrawal of the right index finger from a microswitch lever electrode.

Stimuli were controlled by means of an electronic timing apparatus, readily adjustable to the different modes of stimulus presentation. A chronometer was started at the onset of the first stimulus and was stopped either by the finger response or the termination of the last stimulus by the timing apparatus. The instrumental-conditioning method was used in all groups.

Two rooms were employed: One contained the timing apparatus and *E;* the other contained the withdrawal apparatus and *S*. The finger-withdrawal apparatus consisted of a microswitch upon which *S* rested the right index finger; the other fingers and the hand were prevented from moving by means of a foam-rubber clamp. Movements of the forearm were also prevented by an arm clamp.

To minimize voluntary behavior, *S*s were provided with reading material and were told that the experiment was designed to test the effects of different types of distraction upon reading

rate and comprehension, and that they would be tested later. Subsequent questioning indicated that none of the *Ss* were aware of the conditioning aspects.

*Subjects.*—Twenty male and 20 female students were assigned to forward, backward, simultaneous, and pseudo groups so that each group consisted of five male and five female *Ss*, by closed random-order assignment.

*Procedure.*—A pretest of three tone presentations was given to each *S* to determine the original neutrality of the tone. If *S* made a finger withdrawal to the tone he was eliminated from the experiment. The tones were followed by five presentations of the shock to determine the minimal consistent response threshold and to accustom *S* to this stimulus.

The above trials were then followed by a total of 100 training and test trials. In the forward, backward, and simultaneous groups 80 pairings of the CS and US were presented. These were interspersed with 20 test trials of the CS alone so that one out of every five trials was a test trial. Its position in the block of five trials was randomly determined. The same procedure was followed for the pseudo group, except that shock alone was presented in the training trials.

For both the forward and backward groups, the time interval between the onset of the first and second presented stimuli was .5 sec. The CS and US started at the same time for the simultaneous group. In all groups the tone was presented for .2 sec.; the shock was presented for .2 sec. or until finger withdrawal resulted, if same occured within the .2 sec. All trials were presented during one sitting. Responses which occurred within .7 sec. following the onset of the second stimulus were treated as CR's.

A temporal interval of 10, 15, 20, 25 sec., determined at random, was employed between trials.

## Results

Figure 1 shows the response curves obtained for each group. The curves were constructed by plotting the frequencies of CR's elicited in each group relative to five successive class intervals of four test trials each. The total possible number of CR's per interval would be 40 for each group, i.e., four trials times ten *Ss*.

The forward-conditioning results demonstrate typical negatively accelerated acquisition until the last class

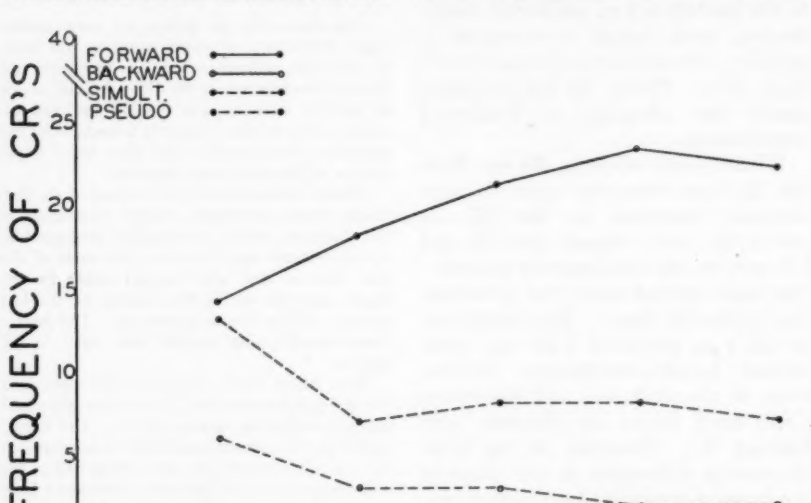


FIG. 1. Conditioning curves showing the frequency of CR's on successive intervals of four test trials for ten *Ss* in each conditioning mode group

TABLE 1

RESULTS OF U TEST FOR EACH INTERGROUP COMBINATION FOR EARLY, LATE, AND TOTAL TEST TRIALS

| Related Groups | Test Trials | | | | | |
|---|---|---|---|---|---|---|
| | 1–8 | | 9–20 | | 1–20 | |
| | *U* | *p* | *U* | *p* | *U* | *p* |
| Forward-Backward | 2.456 | <.01 | 2.721 | <.005 | 2.790 | <.02 |
| Forward-Simultaneous | .907 | —* | 1.927 | <.06 | 1.301 | — |
| Forward-Pseudo | 1.700 | — | 2.343 | <.02 | 2.078 | <.04 |
| Backward-Simultaneous | 1.625 | — | .869 | — | 1.360 | — |
| Backward-Pseudo | 1.209 | — | .756 | — | 1.360 | — |
| Simultaneous-Pseudo | .642 | — | .227 | — | 0.453 | — |

* Dash indicates $p > .06$.

interval. The incidence of CR's in the other groups is infrequent and unstable in successive intervals. The curves of the simultaneous, backward, and pseudo groups may be generally characterized by early acquisition followed by a rapid decrement.

The greatest magnitude of conditioning resulted from the forward sequence with a mean number of 9.8 responses out of a possible 20. The simultaneous method was next with 4.3, followed successively by pseudo with 1.6, and backward with .3 responses.

Large differences in variance among the groups meant that the assumption of homogeneity necessary for the *F* and *t* tests could not be met. The *U* test, developed by Mann and Whitney (**3**), which does not make this assumption, was applied to each intergroup comparison. From the shape of the curves obtained it appeared most meaningful to make comparisons of early trials (1–8) and late trials (9–20) as well as all test trials combined. The results of these comparisons are presented in Table 1. From inspection of Table 1, it may be seen that during test trials 9–20 significantly greater incidence of response occurred in the forward group than in any of the others. Backward, simultaneous, and pseudo results differed insignificantly from each other.

## DISCUSSION

The results of the present investigation appear to be in accord with the findings of related studies. The capacity of the forward method of presentation to yield a greater cumulative occurrence of CR's has also been indicated by other *E*s (**1, 5, 9**). Spooner and Kellogg (**5**), using a similar technique, obtained slightly less forward conditioning than the present study, but the forward curves were remarkably similar. The curve exhibited by the pseudo group appears to be one of early acquisition followed by extinction.

Response curves of the simultaneous and backward groups apparently follow the pseudo curve of early response incidence followed by extinction. These results indicate that responses occurring under the above conditions do not yield learning curves in the conventional sense. Rather, these three curves may show early stimulus generalization to the sudden onset of the CS, since the shock also had such a sudden onset. As this

response to the generalized CS was not reinforced, it apparently became extinguished in later test trials.

## Summary

The present investigation was designed to determine the relative incidence of forward, backward, and simultaneous conditioning when compared with a pseudo-conditioning control. An instrumental finger withdrawal to shock technique was used. Forty college students served as *Ss*—ten assigned to each of the four temporal relationships.

Forward conditioning was found to be significantly more frequent than backward, simultaneous, or pseudo conditioning. The forward curve was a typical negatively accelerated learning curve, whereas the other curves showed early acquisition and later extinction. No significant difference was found to exist between the backward, simultaneous, and pseudo modes of presentation, although each differed significantly from the forward results.

Conditioning seems to occur by presenting the stimuli in a forward sequence only, while stimuli presented under the other conditions give rise to unstable results which may be attributed to early generalization and later extinction.

(Received February 11, 1952)

## References

1. Bernstein, A. L. Temporal factors in the formulation of conditioned eyelid reactions in human subjects. *J. gen. Psychol.*, 1934, **10**, 173–197.
2. Harlow, H. F. Forward conditioning, backward conditioning, and pseudo-conditioning in the goldfish. *J. genet. Psychol.*, 1939, **55**, 49–58.
3. Mann, H. B., & Whitney, D. R. On a test of whether one of two random variables is stochastically larger than the other. *Ann. math. Statist.*, 1947, **18**, 50–54.
4. Porter, J. M., Jr. Backward conditioning of the eyelid response. *J. exp. Psychol.*, 1938, **23**, 403–410.
5. Spooner, A., & Kellogg, W. N. The backward conditioning curve. *Amer. J. Psychol.*, 1947, **3**, 321–334.
6. Switzer, St. C. A. Backward conditioning of the eyelid reflex. *J. exp. Psychol.*, 1930, **13**, 76–97.
7. Wenrick, J. The conundrum of the conditioned response. *Psychol. Rev.*; 1933, **40**, 549–559.
8. Wickens, D. D., & Wickens, C. A study of conditioning in the neonate. *J. exp. Psychol.*, 1940, **26**, 94–102.
9. Wolfle, H. M. Conditioning as a function of the interval between the conditioned and the original stimulus. *J. gen. Psychol.*, 1932, **6**, 80–101.

# HUMAN DISCRIMINATION LEARNING WITH SIMULTANEOUS AND SUCCESSIVE PRESENTATION OF STIMULI

HENRY B. LOESS

*State University of Iowa*

AND

CARL P. DUNCAN

*Northwestern University*

The controversy among supporters of different theories of discrimination learning has recently aroused interest in the question whether it is necessary or even advantageous for the learner to make a comparison between the stimuli to be discriminated. One view is represented by Lashley and Wade (**5**) who maintain that discrimination learning takes place only by comparing the correct stimuli with the incorrect ones. The comparison need not be between immediate sensory impressions; it may be between immediate stimuli and traces of previous ones.

In contrast, the Hull-Spence (**3, 4, 7, 8**) reinforcement theory does not include comparison as an important variable. In this view, differential response is based on the development of habit strength to the positive stimulus and inhibition to the negative one. Since these reaction tendencies are to the specific stimuli, it should make little or no difference whether direct comparison is possible or not.

Lashley and Wade's claim that comparison of stimuli, or comparison of stimuli with traces of previous stimuli, is necessary for discrimination learning cannot, of course, be disproven, for if any learning takes place comparison is assumed to have occurred. However, they also state that amount of opportunity for comparison will affect the *rate* of learning. Thus, when the stimuli are presented successively, learning should be slower or more difficult than when they are presented simultaneously.

Let us assume that the discrimination is a two-choice one only. Simultaneous presentation then means that the stimuli are paired in the usual way; on any one trial both a positive and a negative stimulus confront the subject. Successive presentation is defined as the stimulation of *S* by *either* a positive or a negative stimulus on any one trial and the requirement that *S* make some differential response to the single stimulus before it is removed and another substituted for the next trial.

There are few experiments yet reported which have studied discrimination learning with these techniques. Grice (**2**) appears to have been the first. In his experiment two groups of rats were trained to discriminate between circles of different sizes. For one group (simultaneous) both circles were present every trial; in the other group (successive) only one of the circles was present each trial, the positive or the negative stimulus as determined by a previously prepared random order. In the simultaneous group an error was defined in the usual way; i.e., the response of turning toward the negative stimulus and moving far enough to touch the stimulus card. This could not be applied, of course, in the successive group; with only one stimulus present the animal could not make a differential response of turning away from one stimulus toward another. Instead Grice computed the median latency, measured as the time elapsed between *E*'s opening the starting door and *S*'s touching the stimulus card, for each block of ten trials. Any response to the

positive stimulus with a latency greater than this median, and any response to the negative stimulus with a latency less than the median, was called an error.

Grice found no significant differences between his two groups, either on number of trials to learn or total errors to a criterion of 18 correct responses out of 20 consecutive trials. In addition, there was no significant loss in either group when, after learning was complete, each group was shifted to the other situation and given 20 test trials; positive transfer was nearly perfect. The results therefore failed to support the Lashley and Wade contention that learning should be faster when there is greater opportunity for comparison.

There is one other pertinent experiment. Baker and Lawrence (1) first trained rats on a size discrimination, one group with successive, the other with simultaneous presentation. Both groups were then tested for transposition on a simultaneous discrimination. Significantly greater transposition was obtained with the simultaneous group, an apparent contradiction to the Hull-Spence theory. Objections to the design, however, as pointed out by the authors, concern the use of the simultaneous method in the transposition test, and the increasingly greater proportion of trials followed by reinforcement in the simultaneous group as training proceeded during original learning.

In this controversy over the effect on rate of learning of opportunity to compare the stimuli, one variable seems to have escaped attention; viz., the relative difficulty of the discrimination. This variable should be particularly important. It may be that when the discrimination is relatively easy for *Ss*, it will make little difference whether or not direct comparison is possible. If difficulty is defined as the number of cues, relevant or irrelevant, which *S* must discriminate, then when that number is small, *Ss* may find it just as easy to identify and ignore the irrelevant cues from memory, as from direct observation. When, however, the discrimination is made more difficult by adding more cues, *Ss* working under successive presentation may not be able to recall all of the possible combinations of cues. Such *Ss* might then take longer to learn than *Ss* working under simultaneous presentation where all of the cues are present every trial. In short, we are hypothesizing an interaction between degree of opportunity to compare the stimuli in discrimination learning and difficulty of the discrimination. As the discrimination is made more difficult the method of simultaneous presentation, with its greater opportunity for direct comparison of the positive and negative stimuli, may prove advantageous. This hypothesis is tested in the following experiment.

## Procedure

*Subjects.*—The *Ss* were 140 students drawn from two different undergraduate psychology courses at Northwestern, an elementary course and a more advanced laboratory course. Data were collected on the easier discrimination first, using students from both classes. However, examination of the data revealed that beginning students were poorer learners. Because of this finding, only advanced students were used on the difficult discrimination.

The breakdown of *Ss* into groups was as follows: 20 elementary and 20 advanced students for each method of presentation on the easier discrimination, a total of 80; 30 advanced students for each method of presentation on the difficult discrimination, a total of 60.

*Apparatus.*—The *S* was seated at one end of a small table, *E* at the other end. A masonite panel 30 in. square separated *E* from *S*. In the center of the panel was a sliding window, 6½ in. long by 4 in. high, by means of which *E* presented the stimuli to *S*. On the table in front of *S* were two telegraph keys 4 in. apart and directly below the sliding window. On *E*'s side of the partition there was a switch enabling him to complete an electrical circuit between either of the keys and a buzzer. The buzzer sounded to indicate a correct response to both *E* and *S*. A small light on *E*'s side of the panel flashed if *S* pressed the wrong key.

*Stimuli.*—There were eight stimulus cards in all. Each stimulus consisted of a figure drawn in India ink on a 3½ by 4 in.-piece of heavy cardboard which was then covered with a lantern slide cover glass. Four of the stimuli were printed on white and four on medium-gray cardboard. One card of each color had printed on it one of the following designs: large square (2½ in. to a side), small square (¾ in. to a side), large circle (2½ in. diameter), or small circle (¾ in. diameter). Thus there were three variables:

size of design, shape of design, and color of background, represented in all possible combinations.

Four cards, the large and small circle on white and the large and small square on gray, were arbitrarily called Set A; the other four cards made up Set B. The *S*'s task was to learn to discriminate Set A from Set B; thus he had to learn "circles on white, squares on gray" versus "circles on gray, squares on white." In the present study this defines the easier discrimination.

For the more difficult discrimination the same eight cards were used with the following addition: An equilateral triangle 2 in. to a side was printed on each card. The triangles fell within the large circles and squares and surrounded the small circles and squares. The apex of the triangle pointed up on white cards and down on gray cards. With the triangles printed in the manner described, *S* could make the same discrimination as with the easier problem; i.e., Set A was still circles on white, squares on gray; the triangles were irrelevant cues. The *S* could, however, call Set A circles with triangles pointing up, squares with triangles pointing down, or use all of the cues except size and call Set A circles on white with triangles up, squares on gray with triangles down. Any one of these ways of discriminating Set A from Set B was called correct.

*Instructions to S.*—Since the instructions to *S* were rather long, only the most pertinent points will be presented here. After *E* and *S* were seated, *E* described briefly what *S* would see; i.e., white and gray cards containing large and small circles and squares, and, for *S*s on the difficult discrimination, triangles. All *S*s were told that *E* had divided the cards into two sets, A and B, and that their task was to learn to discriminate the two sets. The *S*s working under simultaneous presentation were told that on each trial two cards, one from each set, would appear in the window. The *S* was to decide which of the two was an A card and press the key immediately below that card, the left key if he thought the card on his left was an A card, the right key if he thought the A card was on his right. If *S*'s choice was correct, *E* sounded a buzzer.

The *S*s working under successive presentation were given the same instruction except for being told that the cards would appear one at a time. For these *S*s a label was printed on each key, the letter "A" on the left key, "B" on the right. The *S* therefore pressed a key following the appearance of each stimulus card, which key depending upon whether he thought it an A or B card. Again, correct responses were reinforced with the buzzer.

All *S*s were told to take as much time as they liked on any one trial but that it was probably to their advantage not to take too much time. The *S*s were also instructed that every few trials *E* might ask them if they had discovered the principle; i.e., which cards belonged in each set.

*Method.*—With each card from Set A to be paired with each card from Set B and including a right-left alteration, there were 32 possible combinations for simultaneous presentation. The order of presentation of the pairs was varied so that the same card never appeared in two successive trials. After 32 trials, the sequence was repeated.

When the cards were presented singly, four different orders were used for the eight cards; thus the sequence was repeated after every 32 trials. Again, the same card never appeared twice in succession.

On each trial the card was allowed to remain in the window until *S* responded. Approximately 12 sec. elapsed between trials. When *S* pressed the key, *E* was informed by the flashing of a light on his side of the panel.

Whenever *S* made three consecutive correct responses, *E* paused and asked him if he had discovered the solution. This procedure was followed until the correct answer was given; i.e., until *S* made it clear he knew which cards were in Set A and which in Set B. Those *S*s who had not solved the problem at the end of 75 trials on the easier discrimination, or after 96 trials on the harder discrimination were not given further practice.

## Results

Five of the six distributions of trials to learn were truncated because a few *S*s were stopped after 75 or 96 trials. The only exception was the distribution for advanced students learning the easy discrimination by the successive method, where the highest score was 66 trials. We shall therefore use the median trials to learn in making group comparisons. Table 1 shows the median trials to learn for all groups.

The significance of the difference between any two of the distributions obtained with different methods of presentation was tested both by the Mann-Whitney *U*-test (6) and by chi square. To test significance by chi square, the procedure was to consider the two distributions being compared as one over-all distribution and to compute this distribution's median.

TABLE 1

MEDIAN TRIALS TO LEARN

| Discrimination | Simultaneous Presentation | Successive Presentation |
|---|---|---|
| Easy | | |
| Elementary students | 36.0 | 29.0 |
| Advanced students | 24.5 | 18.0 |
| Difficult | | |
| Advanced students | 31.5 | 72.0 |

Then a fourfold table was set up in which each *S* was assigned to a cell depending upon whether his score was above or below the median of the overall distribution and whether he had learned under the simultaneous or successive method. The results of these tests of significance are shown in the upper part of Table 2.

We can also compare the two levels of discriminative difficulty when the method of presentation is the same for each level. These comparisons are based only on advanced students since no elementary students were run on the more difficult discrimination. The results of these tests of significance are shown in the lower part of Table 2.

Table 1 shows that median trials to learn the easier discrimination were lower when the stimuli were presented successively. The differences between medians for successive and simultaneous presentations are in the same direction for both elementary and advanced students and are nearly the same magnitude, 7.0 trials for elementary students, 6.5 for advanced. Table 2 shows, however, that neither difference is statistically significant by either test. We must conclude that method of stimulus presentation was not a significant variable with the easy discrimination.

When the discrimination was made more difficult, however, the method of presentation was very important. Table 1 shows that *Ss* working under successive presentation required almost twice as many median trials to learn as those using the simultaneous method. The difference of 30.5 trials is highly significant. Our original hypothesis is supported; the method of presentation becomes important when the discrimination is made more difficult.

We cannot, however, claim that the addition of the triangles made the discrimination more difficult in any absolute sense; in this study it depended upon the method of presentation. This is shown by the comparisons between the easy and difficult discriminations when the method of presentation was the same. Table 1 shows that the median trials to learn were somewhat higher for the difficult than for the easy discrimination when simultaneous presentation was used. The difference of seven trials is not, however, statistically significant. When successive presentation was used, the level of discriminative difficulty was of great importance. The median for

TABLE 2

CHI SQUARES, *U*-SCORE DEVIATES, AND THE CORRESPONDING PROBABILITIES WHEN VARIOUS GROUPS ARE COMPARED*

| Comparison | $\chi^2$ | $p$ | *U*-deviate | $p$ |
|---|---|---|---|---|
| *Simultaneous* vs. *Successive Presentation* | | | | |
| Easy discrimination | | | | |
| Elem. students | 0.1 | >.70 | 1.13 | .13 |
| Adv. students | 0.9 | >.30 | 1.46 | .07 |
| Difficult discrimination | | | | |
| Adv. students | 8.07 | <.01 | 3.38 | <.001 |
| *Easy* vs. *Difficult Discrimination* | | | | |
| Simultaneous presentation | 0.08 | >.80 | 0.93 | >.17 |
| Successive presentation | 10.08 | .001 | 4.60 | <.001 |

* All chi squares corrected for continuity.

the more difficult problem is numerically four times that for the easier discrimination and the difference of 54 is highly significant. Thus, the addition of the triangles to the stimulus cards made the discrimination more difficult only when the stimuli were presented successively.

## Discussion

At the outset, we wish to make it clear that we do not think this experiment is crucial for the continuity-noncontinuity argument. Both of these theoretical positions are primarily based on research with lower animals and may not necessarily apply to the discriminations of human beings. Even if it were claimed that data from human *Ss* are relevant to the theoretical controversy, this experiment does not unequivocally support either position. With the easy discrimination, the method of presentation was unimportant, contrary to noncontinuity theory, while with the difficult discrimination method of presentation was very important, *apparently* contrary to continuity theory. We say apparently because we had originally assumed that our experiment was analogous to that of Grice (2). However, when the paper was nearly completed, Spence (9) published an article in which he distinguished between the situation in which some one cue element is consistently reinforced and another case that he called a pattern discrimination. In this latter case no one cue element is consistently reinforced as compared with the others. Spence therefore predicts that learning a pattern discrimination should be slower with the successive than with the simultaneous method. A review of the discrimination required of our *Ss* shows that it is a pattern; neither size, shape, color, nor orientation was consistently correct. We would expect, then, slower learning with the successive method, and this occurred with the more difficult discrimination. Our results with the easy discrimination, where no difference between the two methods of presentation was found, do not fit this interpretation. However, because of the use of human *Ss*, we do not wish to consider our experiment as necessarily relevant to theories of discrimination learning.

Our main interest lies in the support which the results give to the original hypothesis. When a two-choice discrimination is relatively easy, *Ss* learn as well when a positive and a negative stimulus is presented every trial as they do when only a positive *or* a negative stimulus is shown on each trial. If it is claimed that the discrimination is learned by comparing positive with negative stimuli, then apparently the comparison for the simpler task is as easily made between immediately present stimuli and memory traces of previous ones as between simultaneously present stimuli.

When the discriminative task is more complex, the situation is quite different. Again, if it is assumed that *S* learns by comparing stimuli, successive presentation makes the comparison very difficult. We originally defined discriminative difficulty as the number of both relevant and irrelevant cues. The addition of only one more cue considerably increases the number of combinations of cues, any one of which might be the correct one. Perhaps *S* cannot remember all of the possibilities when the stimuli are presented successively.

As our final point we wish to discuss a methodological problem present in some experiments that have attempted to compare learning with simultaneous and successive stimulus presentation. This problem concerns the definition of a trial, or more generally, the comparability of the response measures used with the different methods of presenting stimuli. It will be recalled that

Grice (2) could not use the same response measure for both methods of presentation; with the simultaneous group he used the standard procedure of counting correct and incorrect responses based on whether the animal turned toward the positive or the negative stimulus. This choice was not open to the successive group, so Grice recorded latencies and defined the correctness of any single response according to whether its latency was above or below this median. Response to the positive stimulus with latency greater than the median, or to the negative stimulus with latency less than the median, was called an error. Baker and Lawrence (1) used the same procedure in their experiment.

It is obvious that the same problem of comparable units of measurement occurs in the present study. Although we were able to use correctness of response as the measure for both methods of presentation, objection still might be raised over the definition of a trial. Any such objection is, of course, important because learning was measured by number of trials and group comparisons were made with that measure.

We defined a trial as the presentation of *either* a pair of stimuli or a single stimulus and a subsequent differential response by *S*. But this means that the minimum number of stimulus presentations ("trials") for *S* to see all eight stimulus cards once each is not the same for both simultaneous and successive methods. With successive presentation at least eight "trials" (assume quotation marks hereafter) are necessary to experience each stimulus once. With simultaneous presentation it would be possible to see all eight cards in only four trials. Should all scores obtained under successive presentation be divided by two? We do not know. If it is insisted that all scores with the successive method should be so reduced, the changes in our findings that would result can probably be determined merely by making the reduction in the appropriate medians in Table 1, even without recomputation of significance tests.

One's position on this matter of the definition of a trial actually would seem to depend upon one's position on the basic problem whether comparison of the stimuli is necessary for discrimination learning. The continuity theorist should find no problem in defining a trial. What is important is the number of reinforcements regardless of method of stimulus presentation. The noncontinuity theorist who maintains that response is to the relation between stimuli and thus emphasizes comparison should consider level of learning as a function of the number of "comparison experiences." One comparison experience can be obtained on the first occurrence of simultaneously presented stimuli whereas it requires a minimum of two stimulus occurrences to have the first comparison experience with the successive method. In any case, it can be seen that the results of the present experiment, showing, as they apparently do, that opportunity for comparison is important at one level of discriminative difficulty but not at another, provide no basis for concluding that one unit of measurement is more "correct" than another.

## Summary

It was hypothesized that the effect of opportunity to compare the positive with the negative stimuli in discrimination learning would vary with the level of difficulty of the discrimination. Two levels of discriminative difficulty were used and for each level half of a group of human *S*s learned under simultaneous presentation of stimuli while the other half learned under successive presentation. It was assumed that successive presentation gave less opportunity for comparison of positive with negative stimuli. Learning was measured in terms of the number of presentations of either a pair of stimuli or a single stimulus before *S* reached a criterion. The following results were obtained:

1. By either method of presenting the stimuli, advanced students learned the easier discrimination more rapidly than elementary students.
2. There was no significant difference between the two methods of stimulus presentation in learning the easier discrimination.
3. Significantly poorer learning of the more difficult discrimination was obtained with successive than with simultaneous presentation of stimuli.
4. The two levels of discriminative difficulty were significantly different in trials to learn only when the successive method was used. They were not significantly different when the simultaneous method was used.
5. It was concluded that the original hypothesis—that the difficulty of

the discrimination determines whether rate of discrimination learning is affected by greater opportunity to compare the stimuli—was supported.

6. The problem of comparable units of measurement in comparisons of discrimination learning under simultaneous and successive presentation was considered.

(Received February 29, 1952)

## References

1. Baker, R. A., & Lawrence, D. H. The differential effects of simultaneous and successive stimuli presentation on transposition. *J. comp. physiol. Psychol.*, 1951, **44**, 378–382.
2. Grice, G. R. Visual discrimination learning with simultaneous and successive presentation of stimuli. *J. comp. physiol. Psychol.*, 1949, **42**, 365–373.
3. Hull, C. L. The problem of stimulus equivalence in behavior theory. *Psychol. Rev.*, 1939, **46**, 9–30.
4. Hull, C. L. *Principles of behavior.* New York: D. Appleton-Century, 1943.
5. Lashley, K. S., & Wade, M. The Pavlovian theory of generalization. *Psychol. Rev.*, 1946, **53**, 72–87.
6. Mann, H. B., & Whitney, D. R. On a test of whether one of two random variables is stochastically larger than the other. *Ann. math. Stat.*, 1947, **18**, 50–60.
7. Spence, K. W. The nature of discrimination learning in animals. *Psychol. Rev.*, 1936, **43**, 427–449.
8. Spence, K. W. The differential response in animals to stimuli varying within a single dimension. *Psychol. Rev.*, 1937, **44**, 430–444.
9. Spence, K. W. The nature of the response in discrimination learning. *Psychol. Rev.*, 1952, **59**, 89–93.

# SOME DETERMINANTS OF RIGIDITY IN DISCRIMINATION-REVERSAL LEARNING [1]

ARNOLD H. BUSS [2]

*State University of Iowa*

In this paper principles and techniques developed in the study of simple learning situations are extended to the problem of rigidity in human conceptual behavior. This approach departs from prevalent views concerning rigidity. Typically, investigators of rigidity have defined it as a trait or characteristic of the individual. Tests have been devised to measure rigidity, and test scores have been correlated with other personality measures (**3, 6, 9, 10**). In the present study the term "rigidity" is used to describe a particular aspect of behavior—inability to learn a new discrimination in response to changing circumstances.

In the experiment to be described, with stimulus materials similar to those of the Vigotsky test (**7**), changes in the degree of rigidity manifested by an individual were studied as a function of schedules of reinforcement. Degree of rigidity was measured in terms of the ability to reverse a discrimination.

## METHOD

There were two series of stimuli. In the prereversal series there was a different schedule for each of four groups: one group received continuous, one partial, and two differential reinforcement. In this paper continuous reinforcement refers to reinforcement of a response every time it occurs; differential reinforcement refers to reinforcement of a response in the presence of one stimulus and nonreinforcement of it in the presence of another stimulus; partial reinforcement refers to reinforcement of a response in the presence of a stimulus on some trials and nonreinforcement of the response in the presence of the stimulus on other trials.

One differential reinforcement group received more nonreinforcements of the designated response than the other differential reinforcement group. It was expected that the additional nonreinforcements would diminish the strength of the designated response, thereby decreasing its resistance to extinction and facilitating reversal of the discrimination. While the number of nonreinforcements varied, the number of reinforcements of the designated response was held constant for all four groups. In the postreversal series all groups were trained on a differential reinforcement schedule.

*Subjects.*—Eighty *S*s from first-year psychology courses were assigned randomly to groups of 20 each, with the condition that the proportion of men to women was the same for all groups.

*Apparatus.*—The stimuli were wooden blocks of varying shapes, colors, and dimensions. The three shapes of the top surfaces were circular, square, and triangular; the three colors were red, yellow, and blue. There were two heights: tall (1.5 in.) and short (.5 in.). There were three top surface areas for each of the shapes, "wide," "medium," and "narrow," respectively: circles (diameter) 1.3 in., 1.0 in., and .5 in.; triangles (altitude × base) 1.5 in. × 3.3 in., .9 in. × 1.8 in., and .5 in. × 1.0 in.; squares (sides) 1.5 in., 1.0 in., and .5 in. Each stimulus differed from every other stimulus with respect to at least one of these properties: shape, color, height, or top surface area.

The apparatus was essentially the same as that used in a previous study (**1**). The *S* was seated at a table in front of a wooden barrier that prevented him from seeing *E*. Near the bottom of the barrier there was a circular aperture, behind which was a shutter consisting of a vertical sliding door. The *E* operated the shutter manually, controlling the exposure time of each stimulus.

[1] Adapted from a dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy, in the Department of Psychology, Indiana University, February, 1952. The writer is indebted to Dr. W. K. Estes for his suggestions and criticisms during the planning and interpretation of the experiment, and to Dr. A. Goss for his help in the execution of the experiment.

[2] Now at the Carter Memorial Hospital, Indianapolis.

There were two telegraph keys in front of *S*, one marked "Vec" and the other "not a Vec." When both keys were depressed, opening of the shutter by *E* exposed the stimulus and closed a circuit which activated a chronoscope. The release of one of *S*'s telegraph keys stopped the chronoscope. After the stimulus had been presented for 5 sec., the shutter closed, but a relay kept the circuit closed and the chronoscope running. Release of one of the telegraph keys resulted in a flash of light from a small bulb that only *E* could see. The presence or absence of this flash of light indicated to *E* which key had been released.

*Procedure.*—The *Ss* read the following instructions:

"You will be shown a series of wooden blocks, one at a time. Some of the blocks are *Vec* and some are *not Vec*. You do not know what a Vec is at the start, but you will learn as you proceed. Decide as best as you can whether a block is a *Vec* or *not a Vec*. After you indicate your choice, you will be told if you were right; you will be told nothing if you were wrong.

"Start by depressing *both keys*. The shutter will open, and you will see a block. If you think the block is a Vec, lift your finger from the key marked *Vec;* if you think the block is not a Vec, lift your finger from the key marked *not a Vec*. The shutter will close. If your choice was correct, you will be told "Right." If your choice was wrong, you will be told nothing. Then the next block will be presented. When the shutter closes, be sure to depress both keys so that you are ready for the next block.

"Are you ready?"

Each stimulus was exposed for 5 sec. Then the shutter was closed whether or not *S* had responded. When the shutter closed, *E* said "Right" if *S*'s response was correct, but *E* said nothing if *S*'s response was incorrect. The *E* recorded the response (*Vec* or *not a Vec*) and the latency, placed the next stimulus in position behind the shutter, then opened the shutter. The interval between successive presentations of stimuli was approximately 10 sec. The second series followed the first series immediately; *Ss* were not told that there were two series.

*Experimental design.*—In both the prereversal and the reversal series the crucial stimulus dimension was *height*. In the first series the *Vec* response was reinforced in the presence of tall stimuli, and the *not a Vec* response was reinforced in the presence of short stimuli. In the second series the *Vec* response was reinforced in the presence of short stimuli, and the *not a Vec* response was reinforced in the presence of tall stimuli. In the first series the order of presentation of stimuli varied from one *S* to the next in a random fashion, but in the second series the order was the same for all *Ss*.

TABLE 1

EXPERIMENTAL DESIGN FOR THE PREREVERSAL SERIES

| Group | *Ss* | Number of Blocks | | Per cent of Correct *Vec* Responses Reinforced |
|---|---|---|---|---|
| | | Tall | Short | |
| I | 20 | 6 | 0 | 100 |
| II | 20 | 6 | 6 | 100 |
| III | 20 | 6 | 12 | 100 |
| IV | 20 | 12 | 0 | 50 |

The experimental design for the prereversal series is shown in Table 1. For Group I (continuous reinforcement) every *Vec* response was reinforced, and every *not a Vec* response was nonreinforced. For Groups II and III (differential reinforcement) *Vec* responses were reinforced in the presence of tall blocks and nonreinforced in the presence of short blocks; *not a Vec* responses were reinforced in the presence of short blocks and nonreinforced in the presence of tall blocks. For Group IV (partial reinforcement) *Vec* responses were reinforced in the presence of only 6 of the 12 tall blocks, and every *not a Vec* response was nonreinforced; the particular six blocks in the presence of which the *Vec* response was reinforced were varied in a random fashion from one *S* to the next.

It should be noted that the number of tall blocks present in the prereversal series does not necessarily equal the number of actual reinforcements of the *Vec* response. Often *not a Vec* responses are made to tall blocks, and they are, of course, not reinforced. Therefore, the number of tall blocks in the first series determines only the number of *possible* reinforcements, i.e., the number of reinforcements received if every tall block is called *Vec*. Similarly, the number of possible nonreinforcements of the *Vec* response in the first series is equal to the number of short blocks. An earlier study (1) has shown that the numbers of actual and possible reinforcements covary.

Groups I, II, and III differed among themselves only in the possible number (0, 6, and 12, respectively) of nonreinforcements of the *Vec* response in the presence of short stimuli. Group IV differed from the first three groups with respect to the percentage of possible reinforcements of the *Vec* response in the presence of tall stimuli: 50% for Group IV and 100% for Groups I, II, and III.

The reversal series was the same for all groups. It consisted of 12 tall blocks and 12 short blocks.

All *Vec* responses to short blocks were reinforced, and all *Vec* responses to tall blocks were nonreinforced. *Not a Vec* responses to short blocks were nonreinforced, and *not a Vec* responses to tall blocks were reinforced. Thus, in the reversal series all *S*s were reinforced for reversing the discrimination learned in the prereversal series.

Only two responses were available to *S:* releasing a key marked *Vec* and releasing a key marked *not a Vec*. The strength of the *Vec* response is inversely proportional to the strength of the *not a Vec* response. Therefore, it is necessary to consider the strength of only one of these responses. The *Vec* response was arbitrarily selected, and only its strength will be examined. Hereafter, when a response is referred to without designation, it is the *Vec* response.

## Results

In this paper the relative frequency of occurrence of a response is used as the primary measure of response tendency. The latency data may be found elsewhere (2).

The frequencies of *Vec* responses in relation to the height of the stimuli during the first series are presented in Table 2. All statistical analyses were made directly on the frequency data, but in the tables the frequencies are expressed in terms of percentages in order to facilitate comparisons.

An analysis of variance of responses to tall stimuli during the first series is presented in Table 3. The lack of significant differences among groups is not unexpected, since all groups received the same number of possible reinforcements of the *Vec* response.

Table 2 indicates that Groups II and III were making more *Vec* responses to tall stimuli than to short stimuli. *t* tests reveal that the differences were significant at the .01 and .001 levels of confidence, respectively, which suggests that both groups learned the height discrimination.

TABLE 2

Per Cent of *Vec* Responses During the First Series in Relation to the Height of the Stimuli

| Group | Tall | Short |
|---|---|---|
| I | 55.0 | |
| II | 62.5 | 38.5 |
| III | 64.5 | 32.9 |
| IV | 60.8 | |

TABLE 3

Analysis of Variance of *Vec* Responses to Tall Stimuli in the First Series

| Source of Variation | *ss* | *df* | *F* |
|---|---|---|---|
| Groups | 14.00 | 3 | 1.09 |
| Error | 322.00 | 76 | |
| Total | 336.00 | 79 | |

In the reversal series the *Vec* response was reinforced in the presence of short stimuli and nonreinforced in the presence of tall stimuli. In order to determine changes in *Vec* response frequencies, the series was divided into halves, as shown in Table 4.

During the first half of the series all groups persisted to about the same extent in making the height discrimination that they had learned in the prereversal series. However, there were differences between groups in over-all response tendency. In descending order of over-all response frequency were the continuous reinforcement, the partial reinforcement, and the two differential reinforcement groups.

The statistical significance of these trends could not be determined by a standard two-way analysis of variance because the frequencies to tall stimuli are not independent of the frequencies to short stimuli, while the group frequencies are independent. Therefore, the method suggested by Kogan (8) was employed. The error estimate is broken down into an inter-*S* error and an intra-*S* error. Computation of the inter-*S* sum of squares involves taking the deviations of the individual *S* means from their own group means, rather than from the grand mean. The intra-*S* sum of squares is the residual term remaining after the groups, tall-short, groups × tall-short interaction, and inter-*S* sums of squares have been subtracted from the total sum

TABLE 4

PER CENT OF *Vec* RESPONSES IN EACH HALF OF THE SECOND SERIES IN RELATION TO THE HEIGHT OF THE STIMULI

| Group | First Half | | | Second Half | | |
|---|---|---|---|---|---|---|
| | Tall | Short | Total | Tall | Short | Total |
| I | 79.2 | 61.7 | 70.4 | 54.2 | 58.3 | 56.3 |
| II | 61.7 | 45.0 | 53.3 | 38.3 | 65.0 | 51.7 |
| III | 65.8 | 38.3 | 52.1 | 36.7 | 54.2 | 45.4 |
| IV | 70.0 | 50.0 | 60.0 | 39.2 | 64.2 | 51.7 |

of squares. The inter-*S* variance is used as an error estimate for the between-groups variance, and the intra-*S* variance is used as an error estimate for the groups × tall-short interaction variance and the tall-short variance.

An analysis of variance of the first-half frequencies is presented in Table 5.

In the second half of the reversal series there were striking changes in response tendencies in all groups. Groups II, III, and IV reversed the height discrimination, responding more frequently to short stimuli than to tall stimuli. Group I failed to reverse the discrimination, making as many responses to tall stimuli as to short stimuli. An analysis of variance of the second-half frequencies is presented in Table 6.

One aspect of the discrimination reversal was the extinction of the tendency to respond *Vec* to tall stimuli. In the first half of the reversal series this response tendency was strong for

TABLE 5

ANALYSIS OF VARIANCE OF *Vec* RESPONSE FREQUENCIES IN THE FIRST HALF OF THE FIRST SERIES

| Source of Variation | *ss* | *df* | *F* |
|---|---|---|---|
| Groups | 30.50 | 3 | 6.30** |
| Tall-Short | 62.20 | 1 | 36.16** |
| Interaction | 1.40 | 3 | .21 |
| Inter-*S* Error | 123.30 | 76 | |
| Intra-*S* Error | 130.40 | 76 | |
| Total | 347.80 | 159 | |

** Significant at the .001 level of confidence.

TABLE 6

ANALYSIS OF VARIANCE OF *Vec* RESPONSE FREQUENCIES IN THE SECOND HALF OF THE SECOND SERIES

| Source of Variation | *ss* | *df* | *F* |
|---|---|---|---|
| Groups | 8.50 | 3 | 1.65 |
| Tall-Short | 48.30 | 1 | 10.93* |
| Interaction | 11.50 | 3 | .85 |
| Inter-*S* Error | 130.10 | 76 | |
| Intra-*S* Error | 336.70 | 76 | |
| Total | 535.10 | 159 | |

* Significant at the .01 level of confidence.

all four groups, with Group I manifesting the strongest response tendency. In the second half of the reversal series this response tendency was partially extinguished by all four groups. However, despite the decrease, the tendency to respond *Vec* to tall stimuli remained strong in Group I—much stronger than in the other three groups.

Another aspect of the discrimination reversal was the increase in *Vec* response frequency to short stimuli. The frequency of responding to short stimuli of Groups II, III, and IV increased considerably in the second half, but that of Group I dropped slightly.

In both halves of the reversal series, as well as in the prereversal series, the differences between Groups II and III were small and statistically insignificant. The additional six possible nonreinforcements in the prereversal series evidently did not affect the rate of learning of either the pre- or the postreversal discrimination.

## DISCUSSION

The first (prereversal) discrimination was learned under all four conditions of reinforcement and nonreinforcement. The second (reversal) discrimination was learned under differential reinforcement, and the speed of reversal proved to be related to

previous reinforcement conditions. The *Ss* reversed the discrimination quite readily after differential or partial reinforcement but did not reverse the discrimination after continuous reinforcement.

Evidently partial reinforcement resulted in less resistance to extinction than continuous reinforcement. The writer could find no studies in which a similar result was obtained, although some experiments have revealed no difference in resistance to extinction after continuous and partial reinforcement (**4, 5, 11**).

There seems to be no obvious explanation for the prolonged extinction after continuous reinforcement. It is true that the reversal series was short (24 trials). However, it was long enough for the differential and partial reinforcement groups to extinguish the discrimination learned in the prereversal series. Nor can failure to learn the discrimination in the first series be invoked as an explanation. In the first half of the second series all four groups were responding more frequently to tall stimuli than to short stimuli, which indicates that they had learned the height discrimination in the first series.

It must be noted that the conditions of extinction in this experiment were not the same as those of a standard extinction experiment. In the latter a response is extinguished in the presence of all stimuli. In this experiment the response was extinguished in the reversal series only in the presence of tall stimuli; it was reinforced in the presence of short stimuli. The effect of this difference in extinction conditions could be tested by performing an experiment in which the response was not reinforced in the presence of *any* stimuli.

The low resistance to extinction manifested by the differential reinforcement groups was not unexpected. It could have been predicted by an S–R reinforcement theory. The differential reinforcement groups built up the weakest response tendency because of the inhibiting effect of the nonreinforcements, and they were better able to reverse the discrimination than the continuous reinforcement group. The latter had built up a strong response tendency in the absence of nonreinforcements. Thus, the rate of learning the new discrimination seems to be inversely related to the strength of the response tendency built up in the prereversal training.

The variables manipulated in this experiment were suggested by research on simple learning situations, principally experiments concerned with extinction and discrimination reversal. The results suggest that the conditions of reinforcement influence (to a surprising extent) the tendency to behave rigidly in a conceptual situation. The utility of principles of reinforcement and extinction in explaining many of the results suggests that further application of learning principles to the study of rigidity in conceptual situations will be fruitful.

## Summary

Rigidity, defined as resistance to shifting from old to new discriminations, was studied in relation to schedules of reinforcement. Eighty college students were divided into four groups of 20 each. The groups varied with respect to the conditions under which they learned the first of two discriminations. One group received continuous reinforcement, one partial reinforcement, and two differential reinforcement. The second discrimination, which was the same for all groups, involved reversing the first discrimination. The stimuli were presented singly, and *S* responded by releasing

one of two telegraph keys. The *E* either said "Right" or remained silent after *S* responded.

The differential and partial reinforcement groups reversed the discrimination but the continuous reinforcement group did not. The strength of a designated response in the reversal series was found to be inversely related to the number of nonreinforcements in the prereversal series. The results were regarded as indicating the utility of the principles of reinforcement and extinction in the study of rigidity in conceptual situations.

(Received February 28, 1952)

## References

1. Buss, A. H. A study of concept formation as a function of reinforcement and stimulus generalization. *J. exp. Psychol.*, 1950, **40**, 494–504.
2. Buss, A. H. Some determinants of rigidity in the learning of successive discriminations. Unpublished Ph.D. dissertation, Indiana Univer., 1952.
3. Cowen, E. L., & Thompson, G. G. Problem solving rigidity and personality structure. *J. abnorm. soc. Psychol.*, 1951, **46**, 165–176.
4. Denny, M. R. The role of secondary reinforcement in a partial reinforcement learning situation. *J. exp. Psychol.*, 1946, **36**, 373–389.
5. Finger, F. W. Retention and subsequent extinction of a simple running response following varying conditions of reinforcement. *J. exp. Psychol.*, 1942, **31**, 120–133.
6. Fisher, S. Patterns of personality rigidity and some of their determinants. *Psychol. Monogr.*, 1950, **64**, No. 1 (Whole No. 307).
7. Hanfmann, E., & Kasanin, J. A method for the study of concept formation. *J. Psychol.*, 1937, **3**, 521–540.
8. Kogan, L. S. Analysis of variance—Repeated measurements. *Psychol. Bull.*, 1948, **45**, 131–143.
9. McAndrew, H. Rigidity and isolation: A study of the deaf and blind. *J. abnorm. soc. Psychol.*, 1948, **43**, 476–494.
10. Rokeach, M. Generalized mental rigidity as a factor in ethnocentrism. *J. abnorm. soc. Psychol.*, 1948, **43**, 259–278.
11. Sheffield, V. F. Extinction as a function of partial reinforcement and distribution of practice. *J. exp. Psychol.*, 1949, **39**, 511–525.

# JOURNAL OF EXPERIMENTAL PSYCHOLOGY

## The issues listed below are available for sale

| Year | Volume | Available Numbers | | | | | | Price per Number | Price per Volume |
|---|---|---|---|---|---|---|---|---|---|
| 1916 | 1 | 1 | 2 | 3 | 4 | 5 | 6 | $1.25 | $7.00 |
| 1917 | 2 | 1 | 2 | – | 4 | 5 | – | $1.25 | $5.00 |
| 1920 | 3 | – | 2 | – | 4 | 5 | – | $1.25 | $3.75 |
| 1921 | 4 | 1 | – | 3 | 4 | 5 | 6 | $1.25 | $6.25 |
| 1922 | 5 | – | – | 3 | 4 | 5 | – | $1.25 | $3.75 |
| 1923 | 6 | – | – | – | – | – | – | —— | —— |
| 1924 | 7 | 1 | 2 | – | – | 5 | – | $1.25 | $3.75 |
| 1925 | 8 | 1 | 2 | 3 | 4 | – | 6 | $1.25 | $6.25 |
| 1926 | 9 | – | 2 | 3 | – | 5 | 6 | $1.25 | $5.00 |
| 1927 | 10 | – | – | 3 | 4 | 5 | 6 | $1.25 | $5.00 |
| 1928 | 11 | 1 | 2 | 3 | 4 | 5 | 6 | $1.25 | $7.00 |
| 1929 | 12 | 1 | 2 | 3 | 4 | 5 | 6 | $1.25 | $7.00 |
| 1930 | 13 | 1 | 2 | 3 | 4 | 5 | 6 | $1.25 | $7.00 |
| 1931 | 14 | 1 | 2 | 3 | 4 | 5 | 6 | $1.25 | $7.00 |
| 1932 | 15 | 1 | 2 | 3 | 4 | 5 | 6 | $1.25 | $7.00 |
| 1933 | 16 | 1 | 2 | 3 | 4 | 5 | 6 | $1.25 | $7.00 |
| 1934 | 17 | 1 | 2 | 3 | 4 | 5 | 6 | $1.25 | $7.00 |
| 1935 | 18 | 1 | 2 | 3 | 4 | 5 | 6 | $1.25 | $7.00 |
| 1936 | 19 | 1 | 2 | 3 | 4 | 5 | 6 | $1.25 | $7.00 |
| 1937 | 20 | 1 | 2 | 3 | 4 | 5 | 6 | $1.25 | $7.00 |
| | 21 | 1 | 2 | 3 | 4 | 5 | 6 | $1.25 | $7.00 |
| 1938 | 22 | 1 | 2 | 3 | 4 | 5 | 6 | $1.25 | $7.00 |
| | 23 | 1 | 2 | 3 | 4 | 5 | 6 | $1.25 | $7.00 |
| 1939 | 24 | 1 | 2 | 3 | 4 | 5 | 6 | $1.25 | $7.00 |
| | 25 | 1 | 2 | 3 | 4 | 5 | 6 | $1.25 | $7.00 |
| 1940 | 26 | 1 | 2 | 3 | 4 | 5 | 6 | $1.25 | $7.00 |
| | 27 | 1 | 2 | 3 | 4 | 5 | – | $1.25 | $6.25 |
| 1941 | 28 | 1 | 2 | 3 | 4 | 5 | 6 | $1.25 | $7.00 |
| | 29 | 1 | 2 | 3 | 4 | 5 | 6 | $1.25 | $7.00 |
| 1942 | 30 | 1 | 2 | 3 | 4 | 5 | 6 | $1.25 | $7.00 |
| | 31 | 1 | 2 | 3 | 4 | 5 | 6 | $1.25 | $7.00 |
| 1943 | 32 | 1 | 2 | 3 | 4 | 5 | 6 | $1.25 | $7.00 |
| | 33 | 1 | 2 | 3 | 4 | 5 | 6 | $1.25 | $7.00 |
| 1944 | 34 | 1 | 2 | 3 | 4 | 5 | 6 | $1.25 | $7.00 |
| 1945 | 35 | 1 | 2 | 3 | 4 | 5 | 6 | $1.25 | $7.00 |
| 1946 | 36 | 1 | 2 | 3 | 4 | 5 | 6 | $1.25 | $7.00 |
| 1947 | 37 | 1 | 2 | 3 | 4 | 5 | 6 | $1.25 | $7.00 |
| 1948 | 38 | 1 | 2 | 3 | 4 | 5 | 6 | $1.25 | $7.00 |
| 1949 | 39 | 1 | 2 | 3 | 4 | 5 | 6 | $1.25 | $7.00 |
| 1950 | 40 | 1 | 2 | 3 | 4 | 5 | 6 | $1.25 | $7.00 |
| 1951 | 41 | 1 | 2 | 3 | 4 | 5 | 6 | $1.25 | $7.00 |
| | 42 | 1 | 2 | 3 | 4 | 5 | 6 | $1.25 | $7.00 |
| 1952 | 43 | By subscription, $14.00 (2 vols.) | | | | | | $1.25 | —— |
| | 44 | | | | | | | | |
| | | List price, Volumes 1 through 42 | | | | | | | $269.00 |
| | | 30% discount | | | | | | | 80.70 |
| | | Net price, Volumes 1 through 42 | | | | | | | $188.30 |

Information about the Journal of Experimental Psychology: as noted in the table, the journal was not published in 1918 and 1919; from 1937 to 1943 it was published at the rate of two volumes per year; from 1944 to 1950, in one volume per year. Beginning in 1951 there are two volumes a year.

Information about prices: when all numbers are available, the price per volume throughout the series is $7.00. For incomplete volumes, the price is $1.25 for each available number. For foreign postage, $.25 per volume should be added. The American Psychological Association gives the following discounts on orders for any one journal:

10% on orders of $ 50.00 and over
20% on orders of $100.00 and over
30% on orders of $150.00 and over

*Address orders and subscriptions to*

## AMERICAN PSYCHOLOGICAL ASSOCIATION

1515 Massachusetts Avenue N.W. Washington 5, D. C.

# Not Just Stimulation On or Off Yes or No

But controllable, measurable amounts of stimulation can be applied with our Model 226B Electronic Stimulator. You can easily use stimulation as the independent variable in experiments with animals and human beings. Write for our bulletin.

**C. J. APPLEGATE & CO.**

1816 Grove Street Boulder, Colorado